

[illegible]

PA (REGC ) UNIV CALIFORNIA.

aa Fischer RL, Ohad N, Kiyosue T, Yadegari R, Margossian L, Harada J,  
pi  
pi Goldberg RB;

DR WPI; 2000-023577/02.  
DR P-PSDB; AAY57037.

PT FIE polynucleotide sequences, used to modulate endosperm development

XX

PS Claim 11; Page 35; 54pp; English.

This is the FIE3 (fertilisation-independent endosperm) polynucleotide sequence. The FIE genes are a new group of plant regulatory genes. The polypeptide encoded by the nucleotide sequence is involved in the control of reproductive development. The FIE sequence shares homology with a polycomb group gene from *Drosophila*. The invention relates to FIE1 and FIE3 polynucleotide sequences (AA378995-237896). The nucleotide sequences can be used to create a transgenic plant containing an expression cassette with an FIE nucleotide sequence operably linked to a plant promoter. The FIE polynucleotide can be used to modulate endosperm development in a plant. The nucleotide sequences can be used to increase or enhance endogenous gene expression e.g. to enhance the expression of polycomb genes which control cell cycling. This can be used to control plant morphology. Enhanced expression can also be used to increase vegetative growth.

AA Sequence 1563 BP; 429 A; 272 C; 368 G; 494 T; 0 other;  
SQ

```
Query Match      100.0%; Score 1563; DB 21; Length 1563;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1563; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
```

[illegible]

Oy	601	aggacacaacccctttaaaccctcaactctgtatctactctgtacgaagatgatcatcgtttcgt	660
Db	601	aggacacaacccctttaaaccctcaactctgtatctactctgtacgaagatgatcatcgtttcgt	660
Oy	661	ttgtgtgaatgttgaactcgttggaatgttattttgtatattgtcttggaagcgtggaagtcacgc	720
Db	661	ttgtgtgaatgttgaactcgttggaatgttattttgtatattgtcttggaagcgtggaagtcacgc	720
Oy	721	tataaatcttcaagtgtggaatttcacctccgtctgatatattcaacccttctgtcattgttgcgt	780
Db	721	tataaatcttcaagtgtggaatttcacctccgtctgatatattcaacccttctgtcattgttgcgt	780
Oy	781	atggacaccaactatlaaataatggtcactgaaagagcttttggacgtacgctcggaaatca	840
Db	781	atggacaccaactatlaaataatggtcactgaaagagcttttggacgtacgctcggaaatca	840
Oy	841	ttcacatgagactgtatccaacaaattccccacaaaattgttccaatttccctgtatt	900
Db	841	ttcacatgagactgtatccaacaaattccccacaaaattgttccaatttccctgtatt	900
Oy	901	acagcttccattctcaataattatgtatgttgaacacgtttgttctgtggaatttatccctc	960
Db	901	acagcttccattctcaataattatgtatgttgaacacgtttgttctgtggaatttatccctc	960
Oy	961	tcaaaagatgttggacaacacggaatctcgtttgttggaaaccaacaactgaaagagaattccct	1020
Db	961	tcaaaagatgttggacaacacggaatctcgtttgttggaaaccaacaactgaaagagaattccct	1020
Oy	1021	ggcgaaggagacttcaagatgttctatattaaagaatacccggttccaaatgttgtatatttggtt	1080
Db	1021	ggcgaaggagacttcaagatgttctatattaaagaatacccggttccaaatgttgtatatttggtt	1080
Oy	1081	atcaaatgtttcttcttgaccctcccatattaaagtctcgtttggacatggaatgaacgaagaagaa	1140
Db	1081	atcaaatgtttcttcttgaccctcccatattaaagtctcgtttggacatggaatgaacgaagaagaa	1140
Oy	1141	gtttatgtctcgtggaatttgaanaagttgcccctcgttttgattatcaaatgttatcacacaat	1200
Db	1141	gtttatgtctcgtggaatttgaanaagttgcccctcgttttgattatcaaatgttatcacacaat	1200
Oy	1201	caatcaaaagtctgtatcatcaggcaaacagcactgtctgtcgaatggaagcagaatttctgt	1260
Db	1201	caatcaaaagtctgtatcatcaggcaaacagcactgtctgtcgaatggaagcagaatttctgt	1260
Oy	1261	tgctcgcggagaaagggaactatattgacgtcttggaacgttatcaacaagatgagcgtctagct	1320
Db	1261	tgctcgcggagaaagggaactatattgacgtcttggaacgttatcaacaagatgagcgtctagct	1320
Oy	1321	tgttaagaatctgtgaatctagagatgcgaagaataatgagatatccattctttatgttaact	1380
Db	1321	tgttaagaatctgtgaatctagagatgcgaagaataatgagatatccattctttatgttaact	1380
Oy	1381	cttgatcatgttgcatactccctgagacacttgagatgagcttcttgaagcctgtttaaagctcaa	1440
Db	1381	cttgatcatgttgcatactccctgagacacttgagatgagcttcttgaagcctgtttaaagctcaa	1440
Oy	1441	ccctctgtaccaacagtgtatacccttctcggagatattgtgtcttatcttcttaatttcaaac	1500
Db	1441	ccctctgtaccaacagtgtatacccttctcggagatattgtgtcttatcttcttaatttcaaac	1500
Oy	1501	acaaggctgtatcccggaagcttatttgcaagaacaacctctctttcataaagctttctagtta	1560
Db	1501	acaaggctgtatcccggaagcttatttgcaagaacaacctctctttcataaagctttctagtta	1560
Oy	1561	ttc 1563	
Db	1561	ttc 1563	
RESULT	2		
AA501160			
AA501160 standard; cDNA; 1626 BP.			

	Query Match	Similarity	40.9%	Score	638.8	DB	22	Length	1626;	
	Best Local	Similarity	73.6%	Pred	No. 2.5e-178;					
	Matches	814;	Conservative	0;	Mismatches	292;	Indels	0;	Gaps	0
OY	196	cgaatcgtcgaaataacttaacttggaggaacgagtcacatgatcttggtctttgattaccatcgaat	255							
Dd	120	caaatgycgaagatgcgcgcgcggttcgacaacccgttgcgcgcgcgttgcaccccgctgaag	179							
OY	256	aagaatcgtaccaaacgttcgaatatgaattcaagaagaaggaaagaacctgttatgcgttt	315							
Dd	180	aagagggagtacacaggttcacccaacaggtcccaagaagaaggaaagtcctccctcatgtccgc	239							
OY	316	gtttccaactccccttgaatgcttgttcttccttcgaatgcttccttcgttacgcgcctggttgaatcgcg	375							

Db	240	gtcttcaacttcacgcgcctccgcgtactcttcaacagatcttcgcacccgtcgcgcgaacccg	299
OY	376	atcacctcgtatacaaatctctcgcagagatggtgcacataccaagcatctgcaatcctaagtcgat	435
Db	300	gttcacttcttatacagtgctcccgaaagggggaatgaatagctgtgtttgcagcttcatcaattgat	359
OY	436	gaagataagagaagatcgcgtttcttacaacgctaaagcttgagcgtgtgcgttaattatggacaac	495
Db	360	gaagataagagcgcgactgcgtttcttacaacgcttcaacgcgtgcgcgtgcgcgtcaacttgaatacccca	419
OY	496	tatctgcgcgtctgagagagatgaanaaggctataatccagatcatctgacgttcaacagctgaacg	555
Db	420	tttgtgtgtgcgcgggaagaaaccaatggtatcaatccgcgtgtatattgtagcgtgcgcataagagag	479
OY	556	attcataagagctctgtggctcgtatgtagatcaatgtagaacgaaatcaagacacacttcta	615
Db	480	atacacagagagtttgttgaaggccaatgggattacataatgaataatcagaatcctaaccattg	539
OY	616	aaacctcaactgtgatactactgctagcaagagatgaatctcgttcctgtttgtgaaatgttgaa	675
Db	540	aaacctaccctccatcgtgtctgtctagcaagaatgaataccgctttagctctcggaaagcttcat	599
OY	676	actggagatagttaattttgacatcttgcctggagcgtcggaggtgcacatgcgtatgaaatcttaag	735
Db	600	acgggaattgttatcccgataatttgcctgagcgtcgggggtccacgcgcaatgaaagttttgaag	659
OY	736	gtgagatttctacccgtctgataattacaacgcttgcctgtagctgtgtgatatggacaccactat	795
Db	660	gtgagatttctacccgtctcgcgaagaatccgattatgcaagtgtgtgtatggaacaataagctt	719
OY	796	aaaataatggtcaatgaaagagtttttgaagctacgctcggagaagatcattcaatcagatgcctat	855
Db	720	aaaatctgtcaatgaagaagagcttctgacacatagttggaagaatcttcatcctgacgcagaat	779
OY	856	gatactcaaaaatccccccacaaaaattgttcccaatctccctgttatcaagcttccatcttcat	915
Db	780	cttccatccgaagttctcccaaccaataacagctgcagtttccagtttctcatagctccagttcat	839
OY	916	acaaattatgtagctgtgtaaccgctgtgtgtgtgataatttcatcctctcacaagaagctgagac	975
Db	840	tcaaactatgttgactgtcaacaggtgcgtgtgtgatttgcctgtctgttcaagaagctgtgac	899
OY	976	aacgagatccgtgtgtgtggacaaccaactgtgaaagagaattctccctgtgcgagggagatctca	1035
Db	900	aacgagattgtgcttttggacaaccaaaatgaagaacaatctccggagagagatcgcgtg	959
OY	1036	gattgtctatataagatcccggttccaagtgtgagatatttggtttatacaagtttctcgt	1095
Db	960	gatatctcttcagaataatccaagttccagaagtgtgacaatttggtttcaaccaatttccgct	1019
OY	1096	gacctcaatttaagttctgttgcgatagagtaatccagaagaagaaggtttatgtctcggat	1155
Db	1020	gactttcatctatcaactcaattgtctatagagaaataggaagggaagatcctacgtatggagag	1079
OY	1156	tgtgaaaagttgcctctctgttttgatbacaagaattatcacacaatcaatcaaaagtctgtta	1215
Db	1080	ctgcagagtagcctcctctgttcttaattgtcaagaattgtctcatcttccaataccaatcccca	1139
OY	1216	atcagagcaaaagacacatgctctgtcgcgatggaagacagatctctgtctgtcgcgaaagacgg	1275
Db	1140	atcagacagacccgcacatgctatcttgatgtgagagacacatcctgagctgtgttgaagatgct	1199
OY	1276	actatagcgctgcgcgcgcgatctac	1301
Db	1200	actatagcgctgcgcgcgcgatctac	1225
RESULT 3			
ID ASO1162			
ASO1162 standard; cDNA; 1428 bp.			
AC	ASO1162;		
XX			





KW seed; pharmaceutical; nutraceutical; polymer; rice; rcallc.pk0001.d2.  
XX  
XX Oryza sativa.  
OS  
XX Key Location/Qualifiers  
FH 25..1188  
FT CDS /tag= a  
FT /product= "Fertilisation-independent endosperm protein"  
XX  
XX WO200116325-A2.  
XX  
XX 08-MAR-2001.  
XX  
XX 30-AUG-2000; 2000WO-US23735.  
XX  
XX 31-AUG-1999; 99US-0151575.  
XX  
XX (DUPO ) DU PONT DE NEMOURS & CO E. I.  
XX (PION-) PIONEER HI-BRED INT INC.  
XX  
XX Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;  
XX Simmons CR, Weng Z;  
XX  
XX WPI: 2001-244407/25.  
XX P-PSDB: AAU00326.  
XX  
XX  
XX New plant fertilization independent endosperm protein for the  
XX production of seed without fertilization is recombinantly produced -  
XX  
XX Claim 7; Page 67; 94pp; English.  
XX  
XX The sequence represents the coding sequence of fertilisation-  
XX independent endosperm clone rcallc.pk0001.d2. Fertilisation-independent  
XX endosperm proteins are plant reproduction proteins necessary for  
XX apomixis, the formation of seeds without fertilisation. Apomixis is  
XX especially useful to agriculture because it eliminates the necessity of  
XX selfing plants to produce genetically identical seed. Such seed is useful  
XX to produce seeds for human and animal food and for commercial milling and  
XX extraction, including the production of useful recombinant products in  
XX the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and  
XX polymers. Embryoless seed production by transgenic plants is less likely  
XX to cause ethical and environmental concern over transgenic plant  
XX production as no gametes are being produced to cross pollinate with other  
XX crops and the seeds cannot germinate.  
XX  
XX Sequence 1387 BP; 389 A; 246 C; 329 G; 422 T; 1 other:  
SQ  
Query Match 38.5%; Score 602; DB 22; Length 1387;  
Best Local Similarity 72.3%; Pred. No. 1.7e-167;  
Matches 782; Conservative 0; Mismatches 300; Indels 0; Gaps 0;  
OY 214 ttagggaacgatacgaatggtgcttgactgcgaataagaacgtaacaagt 273  
DB 70 ttggagatgacaccagtggtggtatcctgtgttcgaagaagaagaataacagagtc 129  
OY 274 acgaatagatccaggaaggaagaacccctgtatgtctgtgttttcaacttcctgat 333  
DB 130 accaatgccttcaagagggaagacgcctctatacgcgcgtatcttcaactcattcgc 189  
OY 334 gctggttcttgatgctctgtaccgctgttgaaatcgatattactctgtacaattgt 393  
DB 190 tccgcactctcaagtttcgcacgtgtgccaatggaagaaggtactctgttatacaatgc 249  
OY 394 ctccgagatggtgcataatcagatgcatactctatgctgcaagaagaagaagtgctg 453  
DB 250 ctgtatgaaggagatattgctgttttgcaatcttatgctgagagataagaatgagct 309  
OY 454 ttctacaagtaagtgtggcggtgtgcaatggaagaaccataatgtgctggctggaaga 513  
DB 310 ttctacacggtgtgtggtgcatgcaatgttgacgaggaaccccaactgtgtgtgctggaaga 369  
OY 514 gtaaaaggtataatccgagtcatactgacgtaacagtgtaaaagattcataagatctgtg 573

DB 370 ctcaatggggtatccgagtcattgtatgctgtgcaagtgagaagaataaagaatttgtt 429  
OY 574 ggtcatgtagatcagtgacgaataatcaggaacacaaccccttaaacctactgtgatt 633  
DB 430 ggccatgtagactcctaataatgaagtaagctcaacataataatccactactgtgta 489  
OY 634 actgtcagaagaagtaaatctgttctgtgtggaatgtgaactggatgataattgtg 693  
DB 490 tcggaagaagaagatgaatcattctgtgtatgaaatgcaactcctggaatgatactgtg 549  
OY 694 atattgtgcagcgtggaagtcacgtctatgaagttcaagttcgaatttcaatccgtct 753  
DB 550 atattgtgcagcgtggaagtcacgtctatgaagttcgaatttcaatccgtct 609  
OY 754 gatattaccgctctgtgtagtctgtgtatggaacacacataataaataatgtaataa 813  
DB 610 gatagtatcgtatctgtgtagtctgtgtatggaacacacataataaataatgtaataa 669  
OY 814 gagtttgcagcgtgtagtctgtgtatggaacacacataataaataatgtaataa 873  
DB 670 gagtttgcagcgtgtagtctgtgtatggaacacacataataaataatgtaataa 729  
OY 874 acaaaattgtccaaatccctgtatcagcttcaatccatcacaataatgtagattgt 933  
DB 730 acaaaattgtccaaatccctgtatcagcttcaatccatcacaataatgtagattgt 789  
OY 934 aacccgtgtgtgtgatttatacctctcaagaagtgtagacaacagatccgtgtgtg 993  
DB 790 aatagtggtgtgtgtgtatttatacctctcaagaagtgtagacaacagatccgtgtg 849  
OY 994 gaaccaacactgaagaagaatctcctgctgaggaaggtcctcagatgtctataatgatac 1053  
DB 850 gaaccaacactgaagaagaatctcctgctgaggaaggtcctcagatgtctataatgatac 909  
OY 1054 ccggttcgaatgtgataattgtgttatacgaagtttctgtgacccatttaagttct 1113  
DB 910 cctatcccgatgtcaatctcgtgtcacaagaatttctgtgacccatttcaacata 969  
OY 1114 gttgcgatagttaatcagaagaagaagtttatactgtcgtgatttgaagaagttccctct 1173  
DB 970 gttacagtggtgaacaggaagaagaagatttgtgttggaattacagcaagtcctccc 1029  
OY 1174 gtttgattacaagaattatcacacaatacaatacaagttcgttaactcggcaacagccatc 1233  
DB 1030 gtaactgtcgaagaagttgtcacatccctcaatcaataatcccaactcaggaactgcaaca 1089  
OY 1234 tctgtcagatggaagcagcagttctgtgtcgtggaaggaagcgtatgagcgtgtgagc 1293  
DB 1090 tcccttgatggaagtactatattgagttgtgtgaggaatggacaatatgagcgtgtgag 1149  
OY 1294 gt 1295  
DB 1150 gt 1151  
RESULT 5  
AAS01172  
ID AAS01172 standard; cDNA; 1498 BP.  
XX  
XX AAS01172;  
XX  
XX 12-SEP-2001 (first entry)  
XX  
XX Fertilisation-independent endosperm cDNA clone p0120.cdeb48r.  
XX  
XX Fertilisation-independent endosperm; plant reproduction; apomixis; ss;  
XX seed; pharmaceutical; nutraceutical; polymer; rice; p0120.cdeb48r.  
XX  
XX Oryza sativa.  
XX  
XX Key Location/Qualifiers  
FH CDS 3..1130

```

FT      /*tag= a
FT      /partial
FT      /product= "Fertilisation-independent endosperm protein"
FT      /transl_except= (pos:837..839, aa:Xaa)
FT      /note= "Contains no start codon. Xaa= unknown"
PN      WO200116325-A2.
XX      08-MAR-2001.
XX      30-AUG-2000: 2000MO-US23735.
XX      31-AUG-1999: 99US-0151575.
XX      (DUPC) DU PONT DE NEMOURS & CO E. I.
XX      (PION-) PIONEER HI-BRED INT INC.
XX      Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;
XX      Simmons CR, Weng Z;
XX      MPI: 2001-24407/25.
XX      P-PSDB; AAU00325.
XX      New plant fertilization independent endosperm protein for the
XX      production of seed without fertilization is recombinantly produced -
XX      Claim 7; Page 65; 94pp; English.
XX      The sequence represents the coding sequence of fertilisation-
XX      independent endosperm clone P0120.cde048r. Fertilisation-independent
XX      endosperm proteins are plant reproduction proteins necessary for
XX      apomixis, the formation of seeds without fertilisation. Apomixis is
XX      especially useful to agriculture because it eliminates the necessity of
XX      selfing plants to produce genetically identical seed. Such seed is useful
XX      to produce seeds for human and animal food and for commercial milling and
XX      extraction, including the production of useful recombinant products in
XX      the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and
XX      polymers. Embryoless seed production by transgenic plants is less likely
XX      to cause ethical and environmental concern over transgenic plant
XX      production as no gametes are being produced to cross pollinate with other
XX      crops and the seeds cannot germinate.
XX      Sequence 1498 BP: 411 A; 304 C; 356 G; 426 T; 1 other;
SQ
Query Match      33.7%; Score 526; DB 22; Length 1498;
Best Local Similarity 68.6%; Pred. No. 5.4e-145;
Matches 724; Conservative 0; Mismatches 331; Indels 0; Gaps 0;
QY      240 ttgtactccatcgaataagaatcgtacaaagtgaagatggaattcagaagggaaga 299
DB      53 ttgttccaaagcagaagagggtacaaagcgtgcacaaagctcaccgggggaagcg 112
QY      300 accctttgatgctgttcttccaacttccttgatgctcgttcttcttgatgctcgttac 359
DB      113 gcaagctcaccgaatcgaatcgaatccttcgaacttcacacactgaagctcttcgcac 172
QY      360 cgcgtggaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatc 419
DB      173 cgtgcggcggaacgcgtgcgaacactgaagctgcctcgaagatgtaatttgcctatcct 232
QY      420 gcaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatc 479
DB      233 gcaagcatatattgatgaggaatgaatggaatcgttctacacactgaatggtggtcgtgga 292
QY      480 cgttaatggaacccatcgttcggtgcgtggaaggaataaagtataaccgagtcattgga 539
DB      293 tctgatgtgcacacgcgtgttagtgcgtgcaggaagcattggaatcgttcggtatcaaa 352
QY      540 cgtcaacagtgaaacgaatcgaatcgaatcgttcggtgcgttcgaatcgaatcgaatcgaatc 599
DB      353 ctgtgcacgtgagaagattacccaagactttctgtgccaatggtgcgaatcgaatcgaatcgaatc 412

```

```

QY      600 caggacacaaccttaaacctcaactgtgttactgctgacgaagatcgtgtcg 659
DB      413 aagaactcaagcatcgaagccttcgttcacatccttcgaagcaagaaagaaacccgttcg 472
QY      660 ttgtggaatgttgaacttgcgtggaatgtaatttgcataatttgcgtgcgtgcgtgcgtgcgt 719
DB      473 gctgtggaatgttgcacacacggagatcgcatttgcatttgcgtgcgtgcgtgcgtgcgtgcgt 532
QY      720 ctatgaatcgaatggtgtggaatcgcgttcgataattacgcgttcgttcgttcgttcgttcgt 779
DB      533 gaatgaatgattgattgattgattgattgattgattgattgattgattgattgattgattgattg 552
QY      780 tatggacacacacatttaaaatgatgacaaaggaaggttttgacgtacgtgcgaagtc 839
DB      593 catgaaacacacgttcaaaatcatgacaaaggaatcgttcgcatatgattgacgaatc 652
QY      840 attcactgactgatgatcacaataattcccaaaaatttgcacaaatttcctgtaatt 899
DB      653 cttaactgactgactgactcaccacaaatttcacaacaaatattgcaatttcgcgtctt 712
QY      900 taagcttccatcacaataattatgtagattgtaacggttcgttcgttcgttcgttcgttcgttc 959
DB      713 gttgctgtatgatactcactatgattgattgattgattgattgattgattgattgattgattgatt 772
QY      960 ctcaagaggtgtggaacagagatcctggtgtgtggaacacacacgtgaagagattctcc 1019
DB      773 gtcnaagaggtgtggaacagagatcctggtgtgtggaacacacacacacacacacacacacacac 832
QY      1020 ttgcgaagaggtcctcaatgcttcatatgaataccggttcacatggtgtgattgatt 1079
DB      833 cggggaaggtgacatgatacttcttcgaagatccttcgtgcgaaggtgattgattgattgattgatt 892
QY      1080 tatcaatgcttctgtgacacacacacacacacacacacacacacacacacacacacacacacacac 1139
DB      893 tatcaatgcttctgtgacacacacacacacacacacacacacacacacacacacacacacacacac 952
QY      1140 gtttattgctgatttgaatgaatgattccttccttccttccttccttccttccttccttccttc 1199
DB      953 agcttctgtcgtggaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1012
QY      1200 tcaatcaatgcttgaatcgaacgaacgaacgaatgctgcgtggaatggaacgaacgaatgctgc 1259
DB      1013 gcaatgcaatcgtcgtatgagcagactgcgtgtcatttgcattgattgattgattgattgattgatt 1072
QY      1260 ttgctgcgaggaagcgaactatattgcgttcgagac 1294
DB      1073 ctgcagcagagattgagcagatattgagcagattgagcagattgagcagattgagcagattgagc 1107

```

RESULT 6

AA01174 standard; cDNA; 1518 BP.

AA01174:

12-SEP-2001 (first entry)

Fertilisation-independent endosperm cDNA clone ses2w.pk0015.b10.

Fertilisation-independent endosperm; plant reproduction; apomixis; ss;

seed; pharmaceutical; nutraceutical; polymer; wheat; ses2w.pk0015.b10.

Triticum aestivum.

Key Location/Qualifiers

FT CDS 88..1224

FT /\*tag= a

FT /product= "Fertilisation-independent endosperm protein"

PN WO200116325-A2.

PD 08-MAR-2001.

PF 30-AUG-2000; 2000WO-US23735.  
 XX  
 PR 31-AUG-1999; 99US-0151575.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E. I.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;  
 PI Simmons CR, Weng Z;  
 XX  
 DR WPI: 2001-24407/25.  
 DR P-PSDB; AAU00327.  
 XX  
 PT New plant fertilization independent endosperm protein for the  
 PT production of seed without fertilization is recombinantly produced -  
 XX  
 PS Claim 7; Page 69; 94pp; English.  
 XX  
 CC The sequence represents the coding sequence of fertilisation-  
 CC independent endosperm clone ses2w.pk0015.b10. Fertilisation-independent  
 CC endosperm proteins are plant reproduction proteins necessary for  
 CC apomixis, the formation of seeds without fertilisation. Apomixis is  
 CC especially useful to agriculture because it eliminates the necessity of  
 CC selling plants to produce genetically identical seed. Such seed is useful  
 CC to produce seeds for human and animal food and for commercial milling and  
 CC extraction, including the production of useful recombinant products in  
 CC the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and  
 CC polymers. Embryoless seed production by transgenic plants is less likely  
 CC to cause ethical and environmental concern over transgenic plant  
 CC production as no gametes are being produced to cross pollinate with other  
 CC crops and the seeds cannot germinate.  
 XX  
 SQ Sequence 1518 BP; 393 A; 337 C; 403 G; 384 T; 1 other;

Query Match 31.9%; Score 498.2; DB 22; Length 1518;  
 Best Local Similarity 65.3%; Pred. No. 9e-137;  
 Matches 731; Conservative 0; Mismatches 388; Indels 0; Gaps 0;

QY 214 ttagggaacagatcaatagttggtcttctgacatcgataagaacatcgataaagt 273  
 DB 115 ttagggtcgagagcggtggtggtctgctgctgacgagcgagagatcaaacg 174  
 QY 274 acgaatagatccaggaaggaacacttctgtatgctgtgttcaacttctgtat 333  
 DB 175 tgcagcaaacacacccagagcgagccgctctacgcatcggtctcaactcagac 234  
 QY 334 gctcgttcttcgatagtcttctacccgtggtggaacatcgatctgtacaattgt 393  
 DB 235 gcccgctactaagaagctcttcgacacgctcgagcaatcggtgacgagcgtgagc 294  
 QY 394 ctgagagatgtgcatatcagcatgtgaacttcctatgctgataagaataagaaagtcg 453  
 DB 295 ctcccccagagtaactgtgctgtctcgaacatcatatgataagcagatcctcagta 354  
 QY 454 ttctacacgtgaagtggtggtggtggtggtgaatggaacccatggttcgctgagga 513  
 DB 355 ttctacactctgagctggtggtggtggtggtggtggtggtggtggtggtggtggtg 414  
 QY 514 gtaaaagtataatcagatctacgacatcgacacagtgaaacagatcatalaagctctg 573  
 DB 415 agcaatgctgcatcctggtcactcaactgtgacacgagaaagtgtttaagatttctt 474  
 QY 574 ggtcatgagatcagtgagcaaatcagacacacactttaaacctcaactgtgatt 633  
 DB 475 ggcacatgtgataataatgataaagaactcaacatcaagccttcgctctcatt 534  
 QY 634 actgctagaagaagatgataatctgctgtgtggaatgtgaacatggtgataatttg 693  
 DB 535 tctgcaagcaagagcagctgtgtgagcaatggaatgcatcacaggtatcgcagcttg 594  
 QY 694 atatctgtgagctgagagctgcatcgctatgaagtcttaagtgatttcatcgcgtc 753

DB 595 attttgctgagagagagctacacgtaatgaagtattgagtggttacttccacccttct 654  
 QY 754 gatattaccgcttctgtgtgtgtatggaacacacatataatataatgtaargaa 813  
 DB 655 gatctaccgaatttgcagctgtgcatgagtaataactgtttaaaatcgtgtaacaa 714  
 QY 814 gagtttgagagctgagagatcattcagcagctgagatgataatcacaatccccc 873  
 DB 715 gaatttggcacaagctgagaaatcccttaacatgagctgacccctcacaatccaa 774  
 QY 874 acaaaatttgcacaatccctgattacagcttccatcacaatataatataatgtagt 933  
 DB 775 acgaatttggtaatttccgctcagacttccgtgtgttccatcacaatgtagt 834  
 QY 934 aaccgtgtgtgtgtgtatttaccctcacaagaagtgtagacaacagatcgtgtgtg 993  
 DB 835 actagtggtgtgtgtgtacttccatccgtcgaaagtgtagacaatgtaattgtctgtg 894  
 QY 994 gaaccacaactgaaagaagattctcctgctgagagagcttcaagtttctattaaaglac 1053  
 DB 895 gaaccacaataaagaagagaggtcccgagaggtagacattgatttctcagaagtac 954  
 QY 1054 ccggttccaatgtgtatatttggttatacaagtttcttctgtgacctccaattagttct 1113  
 DB 955 cctgtgctgtgtgtgtgtatttccatcacaatctcattgatttccattcaatcaa 1014  
 QY 1114 gttcgataggttaatcaagaagaagttatgtctgtgatttgaagaagttccctct 1173  
 DB 1015 ttgcaatagcgcaacggagcaaaatcatalgtgtggaagtgtagagagagcctctct 1074  
 QY 1174 gtttgattacaagttatcacacaatcaatcaatgctgtaatcaggcaaacagccag 1233  
 DB 1075 gtgttaattacccgctgtgtagtaccacaatgcaaaatgccaatggaagatgagtg 1134  
 QY 1234 tctgtcagtagaagcagatcttctgtctgtcgagagcagagctatagcgtcgagac 1293  
 DB 1135 tctgtgtagaagcagatcttctgtctgtcgagagcagagctatagcgtcgagac 1194  
 QY 1294 gtgattaccaagtagcgtctgtgagctgtgtgaagtta 1332  
 DB 1195 gaagtgaacatcaagctgcaaaaatggaagcaactga 1233

RESULT 7  
 AAS01156  
 ID AAS01156 standard; CDNA; 1643 BP.  
 XX  
 AC AAS01156;  
 XX  
 DT 12-SEP-2001 (first entry)  
 DE  
 XX Fertilisation-independent endosperm cDNA clone ccase-b.pk0026.g4.  
 XX  
 KW Fertilisation-independent endosperm; plant reproduction; apomixis; ss;  
 XX seed; pharmaceutical; nutraceutical; polymer; ccase-b.pk0026.g4; corn.  
 OS Zea mays.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 193..1332  
 FT /\*tag= a  
 FT /product= "Fertilisation-independent endosperm protein"  
 XX  
 PN WO200116325-A2.  
 PD 08-MAR-2001.  
 XX  
 PF 30-AUG-2000; 2000WO-US23735.  
 XX  
 PR 31-AUG-1999; 99US-0151575.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E. I.  
 PA (PION-) PIONEER HI-BRED INT INC.

XX Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;  
PI Simmons CR, Weng Z;  
XX WPI: 2001-244407/25.  
DR P-PSDB: AAU00309.  
XX  
PT New plant fertilization independent endosperm protein for the  
PT production of seed without fertilization is recombinantly produced -  
XX  
PS Claim 7; Page 38; 94pp; English.

XX The sequence represents the coding sequence of fertilisation-  
CC independent endosperm clone ccase-b.pk0026.g4. Fertilisation-independent  
CC endosperm proteins are plant reproduction proteins necessary for  
CC apomixis, the formation of seeds without fertilisation. Apomixis is  
CC especially useful to agriculture because it eliminates the necessity of  
CC selling plants to produce genetically identical seed. Such seed is useful  
CC to produce seeds for human and animal food and for commercial milling and  
CC extraction, including the production of useful recombinant products in  
CC the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and  
CC polymers. Embryoleless seed production by transgenic plants is less likely  
CC to cause ethical and environmental concern over transgenic plant  
CC production as no gametes are being produced to cross pollinate with other  
CC crops and the seeds cannot germinate.

XX Sequence 1643 BP; 429 A; 345 C; 456 G; 413 T; 0 other;

Query Match 31.3%; Score 489.4; DB 22; Length 1643;  
Best Local Similarity 66.4%; Prid. No. 3.7e-134;  
Matches 719; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

QY 215 taaggaaacgagtaaatggttggtcttgactccatcgataaagaaatcgtaacaaatga 274  
DB 221 tcgggtgcgagcgcgaggggtcgtctgcgcacgcggaagcggaagtaacaaagcc 280  
QY 275 cgaatagatcaggaaggaagaaacattgtatgtgtgttttaactcttgatg 334  
DB 281 gcgcaagcaacacgagggaggaagcgccgcatatgtatcggttcaacttcagagc 340  
QY 335 ctggtcttcgagatgtcttgtaacgctggtgtaaatcgatcttaccctgacatgtc 394  
DB 341 cgcgtactacgacgtcttcgccacgctcgcgcaacgcgtaacactacacgcgtgcc 400  
QY 395 tcgagatggtgcacatcagcatcgtacatcgtatgctgaagataaagaaagtcgt 454  
DB 401 ttggaatggaatgttcgctcttacaagcttcgttgatgaagataaagatgagtcgt 460  
QY 455 tttaacggttaagtgtggtgtggtgc--gttaatggaacccaatgtgtggtgag 511  
DB 461 tctatccttaagctggtgctgcacalgtgatgtgtcaccacactgctgtgagcag 520  
QY 512 gagttaaaagataatccgagatcattgacgtcaacagtgaaagcattataagtcgt 571  
DB 521 gaacaaatggaatcattcgtggtcattgtgtacagaagaatgtagctaagagcttc 580  
QY 572 tgggtcagggagatcagtgagcaaatcaggaacacactttaaaccttaactgtga 631  
DB 581 ttggcagcagcagcaactaaatgagataaagaactcaacgcttgaaagcttcgcctca 640  
QY 632 ttactgtagcaagatgaatctgtctgttggtgaatgtgaaactgagatgatatt 691  
DB 641 ttctgcgaagcaagatgaatctgttagctatgtaagatgcatacagagatctgalt 700  
QY 692 tgataattgtcgtgagcgtgaggtcatcgatgaagtcctaagtgtgagatttcact 751  
DB 701 tgataattgtcgtgagcgtgaggtcatcgatgaagtcctaagtgtgagatttcact 760  
QY 752 ctgataattacgcttctgagttgtgtatgagaccacttaataatgatgatacga 811  
DB 761 gtgataattacgcttctgagttgtgtatgagaccacttaataatgatgatacga 820

QY 812 aagagtttgcagcagtcgaggaagtaatcacaatgacatgatcatcaaaatcc 871  
DB 821 aagaatttgcgtatcagttgtgacaatcatcatatcatgacgtacactccatcaaaatg 880  
QY 872 ccaaaatttgcacatccctgattattacagctccatcaatcaaatatgtagatt 931  
DB 881 caaaaaatgtccagtttccagcttgatgtgtcgtcagtaacatcaatgtatgt 940  
QY 932 gtaacgcttggttggatattatccctcaagaagtgtagacaagagatcgtgt 991  
DB 941 gtacaagatggtctgtgacttcatcaagaagtgtagacaagatgtgactt 1000  
QY 992 ggaaccacaactgaaaggaattctccgctgagagagcttcaatgttcaatgaat 1051  
DB 1001 ggaaccgagacaagaagaacagagtcctgggagagagcagatcgtatccatgaagt 1060  
QY 1052 accggttcaatgtgtgattattgtttatcaagttcttctgtgacccatgaatt 1111  
DB 1061 atctgtccagatgtgacattgtgttatacaatttcattgtatgttcaatcaat 1120  
QY 1112 ctgttcgataagtaacaggaaggaagttatgtctggtggtttgaaagtgcctc 1171  
DB 1121 agttgcgataagcaacgctgaaagcaaatctacgtgtggaagtaagtaagtcagcc 1180  
QY 1172 ctgtttgattacaagaattatcacacaatcaatcaagtcgtgaatcaggaacagca 1231  
DB 1181 ctgtccatcattgtctgctgtatataatcagcagtgtaatcgcgataagaacagcag 1240  
QY 1232 tgcctgcagatggaagcagcattctgtctgtcgtgagagcagcactatagcgctgg 1291  
DB 1241 tgcctcgtgaggaagcaacatccttgagctgtgtgaagcgcacacatcgtcggtgg 1300  
QY 1292 acg 1294  
DB 1301 atg 1303

RESULT 8  
AAS01170  
ID AAS01170 standard; cDNA; 1686 BP.  
XX  
XX AAS01170;  
AC  
XX  
DT 12-SEP-2001 (first entry)  
XX  
DE Fertilisation-independent endosperm cDNA clone p0107.cbca179r.  
XX  
XX Fertilisation-independent endosperm; plant reproduction; apomixis; ss;  
KW seed; pharmaceutical; nutraceutical; polymer; p0107.cbca179r; corn.  
XX  
XX Zea mays.  
OS  
XX  
FH Key Location/Qualifiers  
FT CDS 394..1533  
FT /tag= a  
FT /product= "Fertilisation-independent endosperm protein"  
XX  
XX WO200116325-A2.  
XX  
XX 08-MAR-2001.  
XX  
XX 30-AUG-2000; 2000WO-US23735.  
XX  
XX 31-AUG-1999; 99US-0151575.  
XX  
XX (DUPLO ) DU PONT DE NEMOURS & CO E I.  
XX (PION-) PIONEER HI-BRED INT INC.  
XX Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;  
PI Simmons CR, Weng Z;  
XX WPI: 2001-244407/25.  
DR P-PSDB: AAU00309.



Sequence 1803 BP; 446 A; 413 C; 491 G; 449 T; 4 other;

1

ID AAS01164 standard; cDNA; 1486 BP

DT 12-SEP-2001 (first entry)

[illegible]

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

08-MAR-2001

31-AUG-1999; 99US-0151575

[illegible]

XX

PT New plant fertilization

CC endosperm proteins are plant reproduction proteins necessary for  
 CC apomixis, the formation of seeds without fertilisation. Apomixis is  
 CC especially useful to agriculture because it eliminates the necessity of  
 CC selling plants to produce genetically identical seed. Such seed is useful  
 CC to produce seeds for human and animal food and for commercial milling and  
 CC extraction, including the production of useful recombinant products in  
 CC the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and  
 CC polymers. Embryoless seed production by transgenic plants is less likely  
 CC to cause ethical and environmental concern over transgenic plant  
 CC production as no gametes are being produced to cross pollinate with other  
 CC crops and the seeds cannot germinate.

CC Sequence 1486 BP; 391 A; 306 C; 380 G; 408 T; 1 other:

DB 927 tggacattggttatacaatttcatgatttcttcaatcatcagttgagatagc 986  
 QY 1126 aatcagaaggaaggttattatgtctggatttgaagaattgcctctctttgatata 1185  
 DB 987 aaccggaagcaaatctactgttgggaagt-acaggtaagcctctcttccatgtct 1045  
 QY 1186 aagttacacatcaatcaaatgttgaatcaagcaagccatctgtctgcatgta 1245  
 DB 1046 cggctgataatcaacagcagtgtaaatccgataagaacaatcgcagttccctgatgta 1105  
 QY 1246 agcacgattctgtctgtctgcagagcaggaactatattgctgagcagc 1294  
 DB 1106 agcacatctctgagctgttgaagcagccacatctgctggttggatg 1154

Query Match 30.2%; Score 472.2; DB 22; Length 1486;  
 Best Local Similarity 68.1%; Pred. No. 4.3e-129;  
 Matches 687; Conservative 0; Mismatches 318; Indels 4; Gaps 2;

QY 289 gaaggagaagaaccttggatgtgttcttcaacttctcagatgctgttcttccat 348  
 DB 147 gaaggagaagccgcgcgtatatagtatcgggttcaacttcatgacgcgcgtactacgac 206  
 QY 349 gcttcgttaccgcgtgtgtaaatcggatctgttataactgtctcggagatgtgccc 408  
 DB 207 gcttcgcacccgttcggcgcaacgcgtgacaacttalcgctgtcttgagatgtgta 266  
 QY 409 atatacgaatcgaatcctatgtctatgataagaagaagagtcgtttacacgtaagt 468  
 DB 267 ttcgctcttctatacagcttactgtatgagataagagtgctgttctactactaagc 326  
 QY 469 tggggtgtggtgc---gttaatgggaaccatattgttcgctgcggaggaataaagata 525  
 DB 327 tggggtcgtgacacgttgaatggtccacacacgtcgtgtggtgagcaggaagaacatggatc 386  
 QY 526 atccgagtcacatgacgtcacacagttgaacgattacataagctctgtggtcatgtagat 585  
 DB 387 atcgggtcatcatattgtgtctacagaagaagttagctaaagcttgttgccatgtgcagc 446  
 QY 586 tcaatgaacgaatcagaagcacacaccttaaaacctcaactgtgtatctgtctagcaag 645  
 DB 447 tcaataaatgataagaactcaacacctgtgaagcctgtcatcatcttctgcaagaag 506  
 QY 646 gatgaatcgttctgttgggaattgtgaactggaatgatatatttggatttggatgta 705  
 DB 507 gatgaatcgttctgttgggaattgtgaactggaatgatatatttggatttggatgta 566  
 QY 706 gctggaagtcacatgcatgaatgaattcgaatgtgatttccacgcgtcatatatttcacgc 765  
 DB 567 gctggaagtcacatgcatgaatgaattcgaatgtgatttccacgcgtcatatatttcacgc 626  
 QY 766 ttgtcgtatgtgtgtatgataccacatataaataatgataagcaagatttggagc 825  
 DB 627 ttgtcgtatgtgtgtatgataccacatataaataatgataagcaagatttggagc 686  
 QY 826 taccgtcgaagaatcttccatcagatgactgtgtatccatcaaatctccccaataattgtc 885  
 DB 687 tatgttgaacaatcatatcatcagatgactgtgtatccatcaaatctccccaataattgtc 746  
 QY 886 caatcccccgtatttacaagcttccatcacaataattatgattgataacgcttggatt 945  
 DB 747 cagttccacgtctgtatgtctgacgtacacttaactatgtgtatgttcaagaagtgctt 806  
 QY 946 ggtgatttatacctcctcaagaagtgatgacaagaagatcctgttctggaacacacactg 1005  
 DB 807 ggtgatttatacctcctcaagaagtgatgacaagaagatcctgttctggaacacacactg 866  
 QY 1006 aagaagaatctccctggcgagaggagcttcaagatgttctataaagaccggttccaag 1065  
 DB 867 aagaagaacagatcctgggaggaagcagatcatccttcagaagaatcatccttcccaaga 926  
 QY 1066 tctgattattgttatacagatttctgtgacctcatattgaagtctgtcgaatagtt 1125

RESULT 11  
 AAS01169  
 ID AAS01169 standard; cDNA; 1642 BP.

AC AAS01169;

DT 12-SEP-2001 (first entry)

DE Fertilisation-independent endosperm cDNA clone p0104.cabbn62r.

KW Fertilisation-independent endosperm; plant reproduction; apomixis; ss;  
 seed; pharmaceutical; nutraceutical; polymer; p0104.cabbn62r; corn.

XX Zea mays.

FT Key Location/Qualifiers  
 FT CDS 29..1414  
 FT /tag= a  
 FT /product= "Fertilisation-independent endosperm protein"

XX WO200116325-A2.

XX 08-MAR-2001.

XX 30-AUG-2000; 2000WO-US23735.

XX 31-AUG-1999; 99US-0151575.

PA (DUPO ) DU PONT DE NEMOURS & CO E. I.

PA (PION-) PIONEER HI-BRED INT INC.

PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;

PI Simmons CR, Weng Z;

XX WPI. 2001-244407/25.

DR P-PSDB; AA000322.

PT New plant fertilization independent endosperm protein for the  
 production of seed without fertilization is recombinantly produced -

PS Claim 7; Page 60; 94pp; English.

XX The sequence represents the coding sequence of fertilisation-  
 CC independent endosperm clone p0104.cabbn62r. Fertilisation-independent  
 CC endosperm proteins are plant reproduction proteins necessary for  
 CC apomixis, the formation of seeds without fertilisation. Apomixis is  
 CC especially useful to agriculture because it eliminates the necessity of  
 CC selling plants to produce genetically identical seed. Such seed is useful  
 CC to produce seeds for human and animal food and for commercial milling and  
 CC extraction, including the production of useful recombinant products in  
 CC the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and  
 CC polymers. Embryoless seed production by transgenic plants is less likely  
 CC to cause ethical and environmental concern over transgenic plant  
 CC production as no gametes are being produced to cross pollinate with other  
 CC crops and the seeds cannot germinate.

XX Sequence 1642 BP; 421 A; 369 C; 435 G; 417 T; 0 other;



Query Match 27.3%; Score 426.4; DB 22; Length 1642;  
 Best Local Similarity 62.7%; Pred. No. 1.6e-115;  
 Matches 680; Conservative 0; Mismatches 401; Indels 3; Gaps 1;

214 ttaggaacagatcaatagttggtcttgaactccatcgaataagaatcgtcaagt 273  
 230 ttaggatgtgaagtagtgaagaggtactcgtctcgttagtcggaagcgaataagccc 289  
 274 aagaatagagttcaagaaggaagaaaccttgtagctgtgttttcaactccattgat 333  
 290 aatagaagaatataccggtgaataacaccgatacgaagccaggtgttcaattcattgac 349  
 334 gctcgttcttcgactgcttcgttacccgctggtgaatcgaatcttactctgtcaaatgt 393  
 350 atgctgactacatagttgcttgcacatgcagttgcaatagttgttaatttaccgattgc 409  
 394 ctccggaagatggtccatatacgaatcgaatcctatgctgtatgaagaataaggaagtcg 453  
 410 ctgagaatgtgtgtttgtctctcctacaataatgatgtatgtaggaataagatgagttca 469  
 454 ttttaccggttaagttggc---gtgtggttaatgtaggaacccatatgttcgagtcgga 510  
 470 ttctacacataagctggtaccatcgaatcgaatgtagtagccagctgttgggcccga 529  
 511 ggaataaagatataacacgactcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 570  
 530 ggaagacatcgtatcattcgtgtcgtatcgtatcgtatcgtatcgtatcgtatcgtatcgt 589  
 571 gtgggtcaatggaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgt 630  
 590 gtggccatgtgttcaatacagatagataagatcgtatcgtatcgtatcgtatcgtatcgtatc 649  
 631 attactctgaacgaagtgatcgtatcgtatcgtatcgtatcgtatcgtatcgtatcgtatc 690  
 650 attctgcgcgaagcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 709  
 691 ttgataattcgtgagctgagctgagctgagctgagctgagctgagctgagctgagctgagctg 750  
 710 ttgactcttgcagaggtgagctgagctgagctgagctgagctgagctgagctgagctgagct 769  
 751 tctgataattcagctgagctgagctgagctgagctgagctgagctgagctgagctgagctgag 810  
 770 accgaaggttggagcttgcagctgagctgagctgagctgagctgagctgagctgagctgag 829  
 811 aagaagtttgcagctgagctgagctgagctgagctgagctgagctgagctgagctgagctgag 870  
 830 aagaatcttgcagctgagctgagctgagctgagctgagctgagctgagctgagctgagctgag 889  
 871 ccacaaatcttgcagctgagctgagctgagctgagctgagctgagctgagctgagctgagctgag 930  
 890 ccaacgaggaatatacagcttgcagctgagctgagctgagctgagctgagctgagctgagctgag 949  
 931 tgaatacgt 990  
 950 tgaatacgt 1009  
 991 tgggaacacacagcgaagaagaatcttccctgagagagagctgagctgagctgagctgagctgag 1050  
 1010 tgggaacacacacagcgaagaagaatcttccctgagagagagctgagctgagctgagctgagctgag 1069  
 1051 taccggttcccaatgt 1110  
 1070 taccggttcccaatgt 1129  
 1111 tctgttgcagataggaatcgaagaagaaggttattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1170  
 1130 cagatgagcaataggaagaagaagaaggttattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1189  
 1171 cctgttttgattacaagaatttatacacaataatcaaatcgaatcgttaataagccaagaacgc 1230  
 1190 ccggtcttaattgacgagctgtgcaacacgaagaatgcaagtcgcgcaataaggaacgcgca 1249

QY 1231 atgtctgtgatggaagcagacatcttctgtgtgagagagacatataagcgctg 1290  
 DB 1250 gttcttctgacgaagcagacatcttctgtgtgagagagacatataagcgctg 1309  
 QY 1291 gacg 1294  
 DB 1310 gacg 1313

## RESULT 12

AAS01158  
 ID AAS01158 standard; cDNA; 1749 BP.

AC AAS01158;

DT 12-SEP-2001 (first entry)

DE Fertilisation-independent endosperm cDNA clone cen3npk0076.b8.

KW Fertilisation-independent endosperm; plant reproduction; apomixis; ss;  
 seed; pharmaceutical; nutraceutical; polymer; cen3npk0076.b8; corn;

KM EST; expressed sequence tag.

OS Zea mays.

Key Location/Qualifiers

FT CDS 131..1516  
 /tag= a  
 /product= "Fertilisation-independent endosperm protein"

PN W0200116325-A2.

PD 08-MAR-2001.

PE 30-AUG-2000; 2000MO-US23735.

PR 31-AUG-1999; 99US-0151575.

XX (DUPO ) DU PONT DE NEMOURS & CO E. I.  
 PA (PION-) PIONEER HI-BRED INT INC.

PI Butler KH, Danilevska O, Miao G, Morgante M, Sakai H;  
 PI Simmons CR, Weng Z;

DR WPI; 2001-244407/25.

DR P-PSDB; AAU00311.

PT New plant fertilization independent endosperm protein for the  
 production of seed without fertilization is recombinantly produced

PS Claim 7; Page 42; 94pp; English.

XX The sequence represents the coding sequence of fertilisation-  
 independent endosperm clone cen3npk0076.b8. Fertilisation-independent  
 endosperm proteins are plant reproduction proteins necessary for  
 apomixis, the formation of seeds without fertilisation. Apomixis is  
 especially useful to agriculture because it eliminates the necessity of  
 selfing plants to produce genetically identical seed. Such seed is useful  
 to produce seeds for human and animal food and for commercial milling and  
 extraction, including the production of useful recombinant products in  
 the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and  
 polymers. Embryoless seed production by transgenic plants is less likely  
 to cause ethical and environmental concern over transgenic plant  
 production as no gametes are being produced to cross pollinate with other  
 crops and the seeds cannot germinate.

XX Sequence 1749 BP; 450 A; 396 C; 451 G; 452 T; 0 other;

Query Match 27.3%; Score 426.4; DB 22; Length 1749;  
 Best Local Similarity 62.7%; Pred. No. 1.6e-115;  
 Matches 680; Conservative 0; Mismatches 401; Indels 3; Gaps 1;

OY	214	ttgaagcaacgctcaataagctctgtggtctcttgactccctacgcgaataagaataacgtaacaagt	273
Db	332	ttaaagatctgtaagtaagtagtgaagagggtactactcgtgtccctagctcggaaagcgaagaaacaagcc	391
OY	274	acgaaatagatctcgaaggaagggaagaaacctcttgatcgtctgtttcaactctctgat	333
Db	392	aatagcagaatataactcgtgaggaaataacccgactctatgcacatcgggtttcaattcttgac	451
OY	334	gctcgtttcttcgatactgctctcgtttacccgcgtgtgaaacatcgatatactctgtacaaatgtc	393
Db	452	atgcgctactactagatgctctcttcgcacgcgcgaatgtaagtagtgaataattacccgatgc	511
OY	394	ctcggaatagtggtccatatacgaatctgcacatctctatgcgtcatgaaagataaagaaagctc	453
Db	512	cttgagaatctgctgttctgtgctctctcaacaaattatgtatgtaagataaagataagataagc	571
OY	454	ttttacaacggttaagtctggc---gtgtgctgttaagtgggaacccataatgtctgcgtctga	510
Db	572	ttctcaacctctaagctctggaaccatcgcatacgaatltgatagcttccgcgtctgtttgtgcgcga	631
OY	511	ggagtaaaaggtacataacatccgagctcatctgacgtcaacacagtgaaacgattatcaagaagctt	570
Db	632	ggaagcaatcgcgatactctgcgtacatctcgtacatattgtgctaccgaaagttgaataagctta	691
OY	571	gttggtcatctggaatcttaagtgaaacgaataacgaacaaacctttaaactactctgtg	630
Db	692	gttggtcattgtgtttcaatacactatgataaagatccatcgtccctcgaaagccatcaactac	751
OY	631	attacgtctgaacgaaggaatgatactcgttctcgtttgtgaatgcttggaaactctgagatctat	690
Db	752	attctcgtcgaagcaagatgtatctatataagctctatgagctatgaaatgctcaactcgtggaattgcac	811
OY	691	ttgatattctctggaagctctggaagctacatcgcataagaaagcttctaagtctgattctatccg	750
Db	812	ttagcttcttgcaaggcgtctggaagcgcatacgtacatgatgtgttagtggttgacttccacct	871
OY	751	tcgtatatttaacgcgtcttctgctatgttgtgtatgtagacacacactatataatagtctaatg	810
Db	872	accgaaggtctggaattttctgcaaaatctgtgagatagacaatactctgttaaaattgtgcaatg	931
OY	811	aaagaagtttggaaagctcgcgcgaaggtcactcaacatcgaatgcagatccatcaaaatc	870
Db	932	aaagaattcttggaataatgtttgtaaaaactatatactatctatgagctcgtgcacatcaaaagctt	991
OY	871	cccccaaaattgtcccaattccctgtatttacaagctctccatccatcaataatatgtatg	930
Db	992	ccaaagcggaataatcacgaagcttccggtcttgactgcgtcgaagtaacactctgacatagttgat	1051
OY	931	tgtaacagtttggtttggttgatcttatactctctcaaaagatgtgtgacaacgaagatccctgttg	990
Db	1052	tgtaacaagatgtgcttgttgactcttccatctctcaataaagatgtlaaagaatgcagtttgcctt	1111
OY	991	tgggaaacccaacactgaaagaagaattctctctcgtcgcgcgaagcgctctcgaatgtctataaga	1055
Db	1112	tgggaaacccaacaaacgaagaacgcgtlaagcctcgtggagggaggtgtatgttcttcaagaag	1171
OY	1051	taccgggttccaatgctgtgataattgtgttatcaaaagtttctcttggaacctccatattga	1110
Db	1172	taccgggtgcgaagaagtgttctatattgtttatgaaatttcaatgtgatttttaactccaac	1231
OY	1111	tcgtgtcgaabaggtlaacgaagaagaaagtttatgtcttggaatttggaaatggatltgcct	1170
Db	1232	cagaatgccaatagccaacaaataaagcgcagatctatgctctggaagtgtgaagtcacgccc	1291
OY	1171	ccgtttttgtttacaagaagtatacgaacatcaactcaaatctcgtatatacggaaacacgc	1230
Db	1292	ccgcgtcttaattgacgcgcgtgtgcaacacgaagatgtcaagctgcgcgataaagacgcgca	1351
OY	1231	atgctgtctgaatgaagaacgacatctctgctgtgtgtggaagcggagactataatgtgcgtcg	1290
Db	1352	gtgtcatctcgaacggaagcagaactccttttgagcgcgcgcgaagaaacgcgcgcgtatctcgtgcgtg	1411
OY	1291	gaag 1294	

Dd	1412 gacg	1415
	RESULT 13	
ID	AAS01165	
	AAS01165 standard; cDNA: 1104 BP.	
XX	AAS01165;	
AC		
DT	12-SEP-2001 (first entry)	
XX		
DE	Fertilisation-independent endosperm cDNA clone pD003.cqpfn34rb.	
KW	Fertilisation-independent endosperm; plant reproduction; apomixis; ss;	
RW	seed; pharmaceutical; nutraceutical; polymer; pD003.cqpfn34rb; corn;	
KW	EST; expressed sequence tag.	
XX		
OS	Zea mays.	
XX		
Key	Location/Qualifiers	
FT	1..823	
CDS	/*tag= a	
FT	/partial	
FT	/note= "Contains no stop codon"	
FT	/product= "Fertilisation-Independent endosperm protein"	
FN		
PN	MO200116325-A2.	
PD		
PA	08-MAR-2001.	
PF	30-AUG-2000; 2000WO-US23735.	
PR	31-AUG-1999; 99US-0151575.	
RA	(DUPO ) DU PONT DE NEMOURS & CO E I.	
PA	(PION-) PIONEER HI-BRED INT INC.	
P1	Butler KH, Danilevskaia O, Miao G, Morgante M, Sakai H;	
P1	Simmons CR, Weng Z;	
DR	WPI: 2001-244407/25.	
XZ	P-PSDB; AAU00318.	
PT	New plant fertilization independent endosperm protein for the	
PT	production of seed without fertilization is recombinantly produced -	
PS	Claim 7; Page 53-54; 94pp; English.	
CC		
CC	The sequence represents the coding sequence of fertilisation-	
CC	independent endosperm clone pD003.cqpfn34rb. Fertilisation-independent	
CC	endosperm proteins are plant reproduction proteins necessary for	
CC	apomixis, the formation of seeds without fertilisation. Apomixis is	
CC	especially useful to agriculture because it eliminates the necessity of	
CC	selfing plants to produce genetically identical seed. Such seed is useful	
CC	to produce seeds for human and animal food and for commercial milling and	
CC	extraction, including the production of useful recombinant products in	
CC	the endosperm e.g.pharmaceutical, nutraceutical, industrial compounds and	
CC	polymers. Embryoless seed production by transgenic plants is less likely	
CC	to cause ethical and environmental concern over transgenic plant	
CC	production as no gametes are being produced to cross pollinate with other	
CC	crops and the seeds cannot germinate.	
SO	Sequence 1104 BP; 307 A; 204 C; 274 G; 319 T; 0 other:	
Oy		
db		

Query Match 26.0%; Score 406.4; DB 22; Length 1104;  
 Best Local Similarity 69.9%; Pred.No.1.le-109;  
 Matches 548; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

```

511 ggagtaaaagtataatccgagtcattcgaactcaacagtgagaagcatcataagaacctt 570
      ||||| ||||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
11 ggaacgaatggagatattcggtgcatactttgtctacaagaanaagttagctaagacctt 70

```

```

/note="Contains no start or stop codon"
/product="Fertilisation-independent endosperm protein"

MO200116325-A2.
08-MAR-2001.
30-AUG-2000; 2000WO-US23735.
31-AUG-1999; 99US-0151575.
(DUPO ) DU PONT DE NEMOURS & CO E I.
(PION-) PIONEER HI-BRED INT INC.
Xx Xx Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;
Xx Xx Simmons CR, Weng Z;
Xx Xx WPI: 2001-244407/25.
DR P-PSDB: AAU00335.
XX XX New plant fertilization independent endosperm protein for the
PT PT production of seed without fertilization is recombinantly produced -
XX XX Claim 7; Page 80; 94pp; English.
XX XX The sequence represents the coding sequence of fertilisation-
CC CC independent endosperm clone p0003 cpged29rb. Fertilisation-independent
CC CC endosperm proteins are plant reproduction proteins necessary for
CC CC apomixis, the formation of seeds without fertilisation. Apomixis is
CC CC especially useful to agriculture because it eliminates the necessity of
CC CC selfing plants to produce genetically identical seed. Such seed is useful
CC CC to produce seeds for human and animal food and for commercial milling and
CC CC extraction, including the production of useful recombinant products in
CC CC the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and
CC CC polymers. Embryoless seed production by transgenic plants is less likely
CC CC to cause ethical and environmental concern over transgenic plants
CC CC production as no gametes are being produced to cross pollinate with other
CC CC crops and the seeds cannot germinate.
Xx Xx Sequence 777 BP; 225 A; 160 C; 168 G; 210 T; 14 other:
SQ

Query Match 19.6%; Score 306.6; DB 22; Length 777;
Best Local Similarity 72.9%; Pred. No. 3e-80;
Matches 422; Conservative 0; Mismatches 154; Indels 3; Gaps 2

Qy 561 taagatcttgtyggtcagtgaattcaagtgaacgaatacagaacacaaccttaaac 620
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 621 tcaactcgtagtactctgtagcaagatgaatatcgtttcgttcttggaattgaaactgg 680
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 681 gatgtcatcttgatcttgctgtagctgtagtgagtcacgcgtatagaagttctaagtgtgga 740
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 741 ttctcatcaatcttcgcaagaagatgatctgtttaagctatgaaatgccatgaccagg 134
Qy 75 ttgcctcatcaatcttcgcaagaagatgatctgtttaagctatgaaatgccatgaccagg 134
Qy 621 tcaactcgtagtactctgtagcaagatgaatatcgtttcgttcttggaattgaaactgg 680
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 681 gatgtcatcttgatcttgctgtagctgtagtgagtcacgcgtatagaagttctaagtgtgga 740
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 741 ttctcatcaatcttcgcaagaagatgatctgtttaagctatgaaatgccatgaccagg 134
Db 135 gactgtcatcttgatcttgctgtagctgtagtgagtcacgcgtatagaagttctaagtgtgga 194
Qy 801 atggtcatgaagaagttttggaagctacgtacgtacgaagaagtcattcaacagagctgagttcc 860
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 861 atccaatctcccccaaatattgccaattcccgtattcaagcagcttcattcatacaa 920
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 921 tatataatctgaacccgttgttgttgatatttatccctcacaagagttgtgacaaga 980
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 375 ctatgttgatgtacaaagatgcttgctgactcatccatcaagaagtgtgacatga 434  
Qy 981 gactctgtgtggagacacacactgaagaatctctctggcggggagcttcagatgt 1040  
Db 435 aactgtgcttgggagacacgaacaaagacagatcctctggggaaagaaatc--gata 492  
Qy 1041 tctatgaatatacccggttccaatgtgtatattctgtttacaaagtttcttgtaact 1100  
Db 493 tctctcagaagactcgtctccagaatgtgacattgtgtttacaaatttctatgtatct 552  
Qy 1101 ccatttaagtctctgtgcgatagtgtaatcagaagaa 1139  
Db 553 tcaattca-atcagctgcgatagtagcaaccgtgaagcaa 590

RESULT 15  
AAS01189  
ID AAS01189 standard: cDNA, 534 BP.  
AC AAS01189;  
AT AAS01189;  
DT 12-SEP-2001 (first entry)  
DE Fertilisation-independent endosperm cDNA clone ses2w.pk0015.b10.  
XX Fertilisation-independent endosperm; plant reproduction; apomixis; ss;  
KW Seed; pharmaceutical; nutraceutical; polymer; soybean; ses2w.pk0015.b10;  
KM EST; expressed sequence tag.  
OS Glycine max.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 2..489  
FT /tag= a  
FT /partial  
FT /product= "Fertilisation-independent endosperm protein"  
FT /note= "Contains no start or stop codon."  
XX  
XX PN WO200116325-A2.  
XX  
XX PD 08-MAR-2001.  
XX PF 30-AUG-2000; 2000MO-US23735.  
XX PR 31-AUG-1999; 99US-0151575.  
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX PA (PION-) PIONEER HI-BRED INT INC.  
XX PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;  
PI Simmons CR, Weng Z;  
XX WPI; 2001-244407/25.  
XX DR P-PSDB; AAU00342.  
XX  
XX PT New plant fertilization independent endosperm protein for the  
XX production of seed without fertilization is recombinantly produced -  
XX  
XX PS Claim 7; Page 88-89; 94pp; English.  
XX  
XX The sequence represents the coding sequence of fertilisation-  
XX independent endosperm clone ses2w.pk0015.b10. Fertilisation-independent  
XX endosperm proteins are plant reproduction proteins necessary for  
XX apomixis, the formation of seeds without fertilization. Apomixis is  
XX especially useful to agriculture because it eliminates the necessity of  
XX selfing plants to produce genetically identical seed. Such seed is useful  
XX to produce seeds for human and animal food and for commercial milling and  
XX extraction, including the production of useful recombinant products in  
XX the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and  
XX polymers. Embryoleless seed production by transgenic plants is less likely  
XX to cause ethical and environmental concern over transgenic plant  
XX production as no gametes are being produced to cross pollinate with other  
XX crops and the seeds cannot germinate.

XX  
SQ Sequence 534 BP; 155 A; 86 C; 118 G; 175 T; 0 other;  
Query Match 19.1%; Score 298.6; DB 22: Length 534;  
Best Local Similarity 74.5%; Pred. No. 5,7e-78;  
Matches 376; Conservative 0; Mismatches 129; Indels 0; Gaps 0;  
Qy 561 taagatcttggggtcagagatcattcaatgaacgaatcagacacaccccttaaac 620  
Db 1 taagatcttggggtcagagatcattcaatgaacgaatcagacacaccccttaaac 60  
Qy 621 tcaacttgatattcagcagcaagaatgaatcgtctgttgggaatgtgaactcg 680  
Db 61 atcaactcgtggtatcggcgaacgaagaatgaatcgtctgttgggaatgtgaactcg 120  
Qy 681 gataatgatttgatattctgtgagctggaggtcagcagatgaagttcgaagtga 740  
Db 121 aatatgatttgatattctgtgagctggaggtcagcagatgaagttcgaagtga 180  
Qy 741 ttctacccgtctgatatattacccgttctgagtgtgtgtgagacacacattaaat 800  
Db 181 ttctacccgtctgatatattacccgttctgagtgtgtgtgagacacacattaaat 240  
Qy 801 atggtcaatgaagaagtttggagcgtacgtcaggaagtcattcacaatgacatgacc 860  
Db 241 atggtcctaagaaggtctctggacatatgtagaanaatcacaatgacagatctcc 300  
Qy 861 atcaaatcccccacaaaattgttccaatccctgtatttacaagcttccattcatacaa 920  
Db 301 ttccaagtttccacaanaatttgtccagtttccctgttacaactgcttcaagtataaa 360  
Qy 921 ttatgtagattgtaacctgtgtgtgtgtgtattatccctccaagaagtgtgacaaga 980  
Db 361 ttatgtgactgttaataaggtgtgtgtgtgtgtattatccctccaagaagtgtgataa 420  
Qy 981 gatccgttctgggagacacacactgaagaagaattctcctgtcgggaaggagcttcaagt 1040  
Db 421 aattactctgttgggaacctcaaaagtgaacgaaccaactccaagggaaggtgtagtgt 480  
Qy 1041 tctatgaatatacccggttccaatg 1065  
Db 481 cctcataaatacccatcttcgattg 505

Search completed: June 25, 2002, 00:17:22  
Job time: 5356 sec





Db	Accession	Gene	Protein	Length
Db	1021	GGCGAGGGAGCTTCAGATGTTCTATTAAAGATACCCGGTCCAAATGTGMAATTTGGTTT		1080
QY	1081	atcaagttcttcttgacctccattaaagtcgtgtgcgaaggtaacaaagaagaag		1140
Db	1081	ATCAAGTTTCTTGAGACCTCCATTAAAGTCTGTGGCATAGAGTAATACAGAAGGAAG		1140
QY	1141	gtttatgtcgggatttgaaaaagttgcccctcgttttggatatacaaaagtttcaacaat		1200
Db	1141	GTTTATGTCTGGGATTTGAAAAAGTTGGCCCTCCTGTTTTGATTACAAAGTTATACACAAAT		1200
QY	1201	caatcaaaagttctgttaataagcaaacagacacatgtctgtcgaatggaacacatcttgct		1260
Db	1201	CAATCAAAAGTCTGTATATAGGCAAAACACCAATGTCTGTGCATGGAGACGACATTTCTTGCT		1260
QY	1261	tgctcgcgaagagagacataatgycgcgtgggaacgtgatlaccgaagtagcgtctcgaagtc		1320
Db	1261	TGCTCGCAGAGAGAGACATATATAGCGGCTGGGACGTGATACCAAGTACCGGCTCGAGTCT		1320
QY	1321	tgtaagaatttgatgaatagaagatgcgaagaataatgagatataccattctttatgttaatt		1380
Db	1321	TGTAGCAATATGATGATTAAGAGATCGACGAAGAAATAGATATCCATCTTTTATGTGAAT		1380
QY	1381	ctgatacagttctactctcccttgaagaccttgaagatgctcttttgaacctgttaacgttcca		1440
Db	1381	CTGATACAGTTCTACTCTCTCCCTTGAAGACCTTGAAGATGCTCTTTGTACCCTTGTTAAAGTCCA		1440
QY	1441	cccttgatacaagatgatacccttctctgagaatttggcttattctcttcttaagttcaataac		1500
Db	1441	CCCTTGATACCAAGATGATATACCTTCTTCTGGAATTTTGCTTATTTCTGTATTCATTAAC		1500
QY	1501	acaagagctgtatccctgggaaccttatttggagagaacacatctcttcaatgaagcttctagta		1560
Db	1501	ACAAGAGCTGTATCCCTGGGACCTTATTGGCAGAACACATCTCTTCAATAAGCTTTCTAGTA		1560
QY	1561	ttc 1563		
Db	1561	TTC 1563		
RESULT 2				
AF129516		1563 bp	mRNA	linear
LOCUS	AF129516			PLN 06-APR-1999
DEFINITION	Arabidopsis thaliana fertilization-independent endosperm protein			
ACCESSION	AF129516			
VERSION	AF129516.1			
KEYWORDS	AF129516.1 GI:4567094			
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
AUTHORS	Spekmatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
TITLE	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
JOURNAL	1 (bases 1 to 1563)			
MEDLINE	Ohad,N., Yedegar,I., Margossian,L., Hannon,M., Michaeli,D.,			
AUTHORS	Harada,J.J., Goldberg,R.B. and Fischer,R.L.			
TITLE	Mutations in PTE, a WD polycomb group gene, allow endosperm			
JOURNAL	development without fertilization			
MEDLINE	Plant Cell 11 (3), 407-416 (1999)			
AUTHORS	2 (bases 1 to 1563)			
TITLE	Ohad,N., Yedegar,I., Margossian,L., Hannon,M., Michaeli,D.,			
JOURNAL	Harada,J.J., Goldberg,R.B. and Fischer,R.L.			
MEDLINE	Direct Submission			
AUTHORS	Submitted (18-FEB-1999) Plant and Microbial Biology, University of			
TITLE	California, 111 Koshland Hall, Berkeley, CA 94720-3102, USA			
JOURNAL	Location/Qualifiers			
MEDLINE	1. 1563			
AUTHORS	/organism="Arabidopsis thaliana"			
TITLE	/cultivar="Landsberg erecta"			
JOURNAL	/db_xref="taxon:3702"			
MEDLINE	/chromosome="3"			
AUTHORS	/map="between EMB29 and AXR2"			
TITLE	1. 1563			
JOURNAL	gene			





**TITLE** Simmons,C.R. and Weng,Z.  
**JOURNAL** Plant reproduction proteins  
Patent: WO 016325-A 9 08-MAR-2001;  
ST PATENT OFFICE AND CONVENTION RIGHTS : DISCOVER HI-Bred

BASE COUNT	455	a	355	c	404	g	412	t
ORIGIN								

Query Match	40.9%;	Score 638.8;	DB 6;	Length 1626;
Best Local Similarity	73.6%;	Pred. No. 2.4e-157;		
Matches 814;	Conservative 0;	Mismatches 292;	Indels 0;	Gaps 0;

Qy	1036	gagttcatttaagaatcaccgggtccccaatgtagatattgtttatcaaatcttcgt	1095
Qy	1036	gagttcatttaagaatcaccgggtccccaatgtagatattgtttatcaaatcttcgt	1095
Db	960	GATATCTTTCAGAAATATCCAGTTCOCAGAGTGACATTTGGTTCATCAAAATTTCTCGT	1019
Qy	1036	gaccctcaattaaagttctgttcgtcgatagtaatacgaagaagaagtttaagtcggagat	1155
Db	1020	GACTTTCATTATTCATCTCAATTCGTAATAGGAAATAGGAAAGGAAAGATATTCGTATAGGGAG	1079
Qy	1156	ttgaagaattcccccctcgttttgaatacgaagttatcacacaatcaatcgaagtcgtta	1215
Db	1080	CTGCAGAGTAGGCCCTCCCTGTTCTTAATTCGAAAGTTGTCATTCCTCCATTCAAAATCCCA	1139
Qy	1216	atcaaggaacaacagccatgctcgtcgatggaatgaaacgaagatcttcgtctggttggaagaagg	1275
Db	1140	ATCAACACAGACCGCATGCAATTTATGGAGCACAAATCTGAGCTGCTGAGGATGCT	1199
Qy	1276	actatataggcgttggaacgtgataac	1301
Db	1200	ACTATATGGCGCTGGATGCATTTAC	1225

[illegible]

REFERENCE	1 (bases 1 to 1428)
AUTHORS	Butler,K.H., Danilevskaya,O., Miao,G.H., Morgante,M., Sakai,H., Simmons,C.R. and Weng,Z.
TITLE	Plant reproduction proteins
JOURNAL	Patent: WO 0116325-A 13-08-MAR-2001; <i>Genetic Recombination and Compound (US) ; Pioneer Hi-Bred International Inc.; Des Moines, IA; USA</i>

BASE COUNT	ORIGIN
387 a	284 c 326 g 431 t

Query Match	39.0%	Score 609	DB 6	Length 1428
Best Local Similarity	71.9%	Pred. No. 1.7e-149		
Matches 795: Conservative	0	Mismatches 310	Indels 0	Gaps 0

Db 342 TTGGTGGCTGAGAGACTTAATGAAATTAATTCGATTATGATACCTGGCAATGAGAGATA 401  
 QY 559 cataaagctctgtggtgcaatgagatcaglyaacgaaatcagacacaaacttaaaa 618  
 Db 402 TACAAAGCTTTGTGGGTGATGGGATTAATTAACGAATTCGAACTCCAGCGCTGAAA 461  
 QY 619 cctcaactgtgtactctctagcaagatgaaatcgttcctgttcttggtggaatgttgaac 678  
 Db 462 CCATCCTCTGTTGTGTGACAGAACAAAGATGAAATCTGTGCGCTGTGAGATTAATCACT 521  
 QY 679 gggatattgatttgaattctgtgagctgagagctcagatcagatcaagttcaagtgtg 738  
 Db 522 GGGATATGATTTTGTATTTTCTGTGCTGGTGCTGATCGCATCGCAATGAGTTCTAGTGTG 581  
 QY 739 gatttcacgtcgtcgtatattacgcgttctgtagtctgtgtagttagttagttagttagt 798  
 Db 582 GACTTCATCTCTTGTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 641  
 QY 799 atatgtaataagaagattttggaagctgctgagaaagtaatcattcaatgagactgtat 858  
 Db 642 ATCTGGTCAATGAAAGAAATTTTGGACATATGAGAGAAATCTTTTACTTGGACTGATCT 701  
 QY 859 ccatcaaaatccccaacaaatgtgcacatccctgtattacagctcttcacattacata 918  
 Db 702 CCTCTTAATTTCCCAAAATATGTCAGTTCCCAATTTATTTGCTTCAAGTGCATACG 761  
 QY 919 aattatgtagatgttaacggtgtgttgtagtattacccctcacaagaagtgtagaacaac 978  
 Db 762 AACTATGTGATGTCACACCGGTGATGGTATGATTTATGCTCCCAAGACCGTTGATTAAT 821  
 QY 979 gagatccgtgtgtggaac 1038  
 Db 822 GAACCTGTATTTATGGAACCAAAATGAAAGAACACATCTCTGGAAGAGGTATACAGTCGAC 881  
 QY 1039 gtcttaataagatcccggtgtgtaagtgtagatattgttttaagaattttcttgtagc 1098  
 Db 882 ATTTCTCAAAAGATATCTGTTCCCGAATGCGAATTTGGTTTTCAAATTTTCTCTCGAT 941  
 QY 1099 ctccatttaagtcgtgtgtagatagtagtaacaggaagaagaagttatgtctgtagttag 1158  
 Db 942 TTCCATTCAACAGACACACACAGTAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1001  
 QY 1159 aaaagtgccctcctgtgtttgtattacaagaattatacacaatcaatcaagaattctgaac 1218  
 Db 1002 CAAGCCAAACCCCGGTCTTCATGTCAGATTAATTCATATTCAGTCAAAATTCCTCAAT 1061  
 QY 1219 aggaacaacagcagatgtctgtgtagaagcaagcaagattctgtgctgtagaagcagagact 1278  
 Db 1062 AGATTGACATGCCATGCTCATATGATGAGGAGACGATTTCTGCTGTGTGTGAGATGGAAC 1121  
 QY 1279 atatgagcgtggaagctgattacca 1303  
 Db 1122 ATATGGCATGGATGTGTGATGCA 1146

RESULT 5  
 AX090446 1387 bp DNA linear PAT 21-MAR-2001  
 LOCUS AX090446  
 DEFINITION Sequence 35 from Patent WO0116325.  
 ACCESSION AX090446  
 VERSION AX090446.1 GI:13444285  
 KEYWORDS  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.  
 1 (bases 1 to 1387)  
 REFERENCE Butler, K.H., Danilevskaya, O., Miao, G.H., Morgante, M., Sakai, H.,  
 AUTHORS Simmons, C.R. and Weng, Z.  
 TITLE Plant reproduction proteins

JOURNAL Patent: WO 0116325-A 35 08-MAR-2001;  
 E.I. DU PONT DE NEMOURS AND COMPANY (US) ; Pioneer Hi-Bred  
 International, Inc. (US)  
 FEATURES Location/Qualifiers  
 source 1..1387  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 BASE COUNT 389 a 246 c 329 g 422 t 1 others  
 ORIGIN  
 Query Match 38.5%; Score 602; DB 6; Length 1387;  
 Best Local Similarity 72.3%; Pred. No. 1.2e-147;  
 Matches 782; Conservative 0; Mismatches 300; Indels 0; Gaps 0;

QY 214 tttaggaacagatcagatagttgtgtcttcttaccatccatcgaataagaatcagatcagatg 273  
 Db 70 TTGGATGTGATCCAGAGTGGGATCCTTGCTTGTTCGAAGAAGAGATATACAGATC 129  
 QY 274 acgaataagatcgaagaaggaaacacttgatgtagtctgttcaacttcctgtat 333  
 Db 130 ACCAATCGCCTTCAAGAGGAAAGCCGCTCTATACCGCTCATTTTCAACTCATCGAC 189  
 QY 334 gctcgttcttgatgtcttctgttaccgctggtggaatcgtgattactctgtacaattgt 393  
 Db 190 TCCCGCTACTTCAAGCTTTTGCACACTGTGGGCGGCAATAGGCTTACTGTTATCAATGC 249  
 QY 394 ctcggaagatgtccatcatalcagatcgaatccatcctatcgaagaagaagctg 453  
 Db 250 CTGTGATGAAGGAGATATGCTGTTTTCATCTTATGCGATGAGATATGAGATGAGTCT 309  
 QY 454 tttaacacggttaagtcgtgtgtggttaagtgaaggaaccatagctgtcgtgtgaaga 513  
 Db 310 TTTTACACCTGTGGTGGGATGCAATGTTGACGAGCCACCTGTTGCTGTGGCTGAGGA 369  
 QY 514 gtaaaagtataatccgagtcatttgtagtcaacagtgtaaacgattcataagagctgtgt 573  
 Db 370 CTCATGCGGTAAATCCGATCATTTGATGTCGAGTAGAATACATTAAGATGTTGTT 429  
 QY 574 gttcagtagaattcagtgaaacgaacacacacacacacacacacacacacacacacacacac 633  
 Db 430 GGCATGAGACATCCATTAATGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 489  
 QY 634 actgttagcaagatgaatcgt 693  
 Db 490 TCGGCAAGCAAAAGATGATCTATGCGGTATGGAATGCTCATGATGATGATGATGATG 549  
 QY 694 atattgtgtgagctgagtagtctgtatgaagatcgtatgaagttcgtatcgttccgttct 753  
 Db 550 ATATTTGCTGGAGGCGGGGACATCGTAATGAAGTCTTAAGTGTGATTTTCATCCATCG 609  
 QY 754 gatatattaccgttctgttagtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 813  
 Db 610 GATATGATGATGATTTGTGATTTGTGATTTGTGATTTGTGATTTGTGATTTGTGATTTGT 669  
 QY 814 gaattttagacgtacgtcgaagaatcattcacatgtagttagttagttagttagttagttagt 873  
 Db 670 GAGTTGTGACATATGAGAAATCATCATGACATGACATGACATGACATGACATGACATGAC 729  
 QY 874 acaaaattgtccaattccctgtattttagagcttccattcatatacaaatatgttagattgt 933  
 Db 730 ACAAAATTTGTCCAGTTCTCTGTTTACACGCTTCACTGATGATTAATTAATTAATTAAT 789  
 QY 934 aaccgt 993  
 Db 790 AATAGGAGTGGGTGATTTTATTCCTCTCAAGAGTTCATATGAAATTAATTTATCTTGTGG 849  
 QY 994 gaaccaaacactgaagaagatctcctgtgtagagagcttcagatgttctataatgaatagac 1053  
 Db 850 GAACCTTAAGTGAACGAACAACTCCAGGAGAGGAGTGTGATGATGATGATGATGATGATG 909  
 QY 1054 csgttccaatgtgtgatatattgtgttatacaagtttctgtgtaccatcattcaagttct 1113

Db 910 CCTATTCCTCCGATTCGCAATATCTGGTTCAATCAAGTTTCTTGGATTCATTTCAACATA 969  
 Oy 1114 GTTGAGATGATCAATCAAGGAAAGGTTATGTCGGATTTGAAAGTtgcctcct 1173  
 Db 970 GTTACAGTGGGTAACAGGAAAGGAAATTTTGTGGGATTTACAGCAAGTCTCC 1029  
 Oy 1174 GTTGTATACAAGTATCAACAATCAATCAAGTCTGTATCAAGGAAAGGAAAGT 1233  
 Db 1030 GTACTGTGCTGCAAAAGTGTCAATCTCAATCAATCAATCAATCAATCAATCAATCA 1089  
 Oy 1234 tctgtcgaatgaaagcaagcattctgtctgtcgaagagagcaataatgagcgtgagc 1293  
 Db 1090 TCCCTTGTGATGGAAGTACTATTTAGTCTGTGAGGATGGGCAATATGGCGTTGGAT 1149  
 Oy 1294 gt 1295  
 Db 1150 GT 1151

RESULT 6  
 AX090444 1498 bp DNA linear PAT 21-MAR-2001

LOCUS AX090444  
 DEFINITION Sequence 33 from Patent WO0116325.  
 ACCESSION AX090444  
 VERSION AX090444.1 GI:13444284  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Oryza sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
 AUTHORS Butler,K.H., Danilevskaia,O., Miao,G.H., Morgante,M., Sakai,H.,  
 1 (bases 1 to 1498)  
 Simonon,C.R. and Weng,Z.  
 Plant reproduction proteins  
 Patent: WO 0116325-A 33 08-MAR-2001;  
 E.I. DU PONT DE NEMOURS AND COMPANY (US) ; Pioneer Hi-Bred  
 International, Inc. (US)  
 TITLE  
 JOURNAL

FEATURES  
 source  
 1..1498  
 Location/Qualifiers

unsure /organism="Oryza sativa"  
 BASE COUNT 839  
 411 a 304 c 356 g 426 t 1 others  
 ORIGIN

Query Match 33.7% Score 526; DB 6; Length 1498;  
 Best Local Similarity 68.6% Pred. No. 1.2e-127;  
 Matches 724; Conservative 0; Mismatches 331; Indels 0; Gaps 0;

Oy 240 tttagctcatcgaataagaaatcgtacaagtgacgaatgattcgaagaaagaa 299  
 Db 53 TTTGGTTCAAGCAAGAGAGAGAGAGTACAGGCGTCGCAAGCTCACCGAGGGAACG 112  
 Oy 300 accttgaatgctgttcttcaacttccatgagctgttcttcgatgtctgttac 359  
 Db 113 GCAGCTTACAGCGCATCGGATTCACACTTCCGACTTCACTACGAGGCTTGGCCAC 172  
 Oy 360 cgtctgtggaatcgtgattactctgtacaatgtctcggagatggtgcataatcgcatt 419  
 Db 173 CGTCGGCGGCAACCGCGTGCACACCTACAGCTGCTCAAGATGGTAATTTGGTATCT 232  
 Oy 420 gaaatccatagtctgtaagaagaagagtcgtttacaagaagtaagtggcgctgtg 479  
 Db 233 GCAAGCATATATTGATGAGGATTAAGATGAATCGTTCTACACACTAGTTGGGCTGTGA 292  
 Oy 480 cgttaatggaacacatattgtcgtgagagagtaaaaggtataatccgagttatga 539  
 Db 293 TCTTGATGCAACACCGCTTGTAGTGTCTCAGGAACATGGGATCATTCGGGTATCAA 352  
 Oy 540 cgtcaacagtgaacgattcataagaagcttctgtgtgtcattgagattcagtgaaagaaat 599

Db 353 CTGTCCACTGGAAGTACTCAACAGACTTTTGTGGCATGGCCATTCATTAACAGAGAT 412  
 Oy 600 caggaacacacctttaaaccctcaactgtgattactctgtagcaagaaatgaatctgttcg 659  
 Db 413 AAGAACTCAAGATTAAGGCTTGGCTCATATCTCTCAAGCAAGATGATCTGTGATG 472  
 Oy 660 ttgtgaaatgttgaactcgtggaatgtatattgtatattgttgagagtgagtcacg 719  
 Db 473 GGTGGGAATGTTCAACACAGGAGATCTGCAATTTGATTTTGTGGAGCAGGAGGTCACCG 532  
 Oy 720 ctatgaatctcaagtgtgagatttcaatccgtctgaatatcttcacgcttgcattgtg 779  
 Db 533 GAATGAAGTATGAGTGTGACTTCCACCATCTGTATCTACCCGATGCAAGTTGTGG 592  
 Oy 780 tatgacaccactataataatgltcaatgaagaagtlttgagcgtacgtcgtgagaatc 839  
 Db 593 CATGGATACACTGTTAAATATGTCATGAAGAAATTCGGCCATATGTGAGCAATC 652  
 Oy 840 atcacaatgactgtatgcatcaatcaaatcccaacaaatltgtccattccctgtatc 899  
 Db 653 CTTTACATGAGACTGACCTTCCATCAAAATTTCCAAACAAATATGCAATTCGGTCTT 712  
 Oy 900 tacaagcttcaatcacaatattgtatgattgtaacgctgtgtgtgtgtattatcct 959  
 Db 713 GGTGGTGTAGTACATTTCTACATATGTGATTTACTAGATGGCTGTGACTTCAATCT 772  
 Oy 960 ctcaagaagtgtgacaacagagatccgtgtgtggaacacaaactgaagaagattctcc 1019  
 Db 773 GTCAAAGAGTGTGACAAATGAATTTGGCTGGAGCCAAACAAACAAAGCAATTC 832  
 Oy 1020 tggcgaagagcttccagatgtctcattgaataaccggttccaaatgtgtatattgtt 1079  
 Db 833 CGGGAGAGTACATGATTTCTTCCAGAGTATCTGTCGCAAGATGATATCTGGT 892  
 Oy 1080 tatcaagttctctgtgaccttcaattgaagtcgtgtgagatgaatcaagaagaa 1139  
 Db 893 TATCAAAATCTCATGCGCAATTTTCACTTCAATCAATTTGGCAATGAGCAACCGTAAGAAA 952  
 Oy 1140 gtttactgtcgtgatttgaagaagltgcctcctgttcttgaatacaagattatcaacaa 1199  
 Db 953 AGTTTTGTCTGGAAGTACAGTCCAGTCCGCTTTTAACTCGTGGTACATTAATCC 1012  
 Oy 1200 tcaatcaaatctgttaatcaggaacaaagccatgtctgtgagtgaagcagattctgc 1259  
 Db 1013 GCATGCAAAATCTCGATTAAGGACAGACTGCCGTGTCAATTTGATGGAAGCAACATCTTGC 1072  
 Oy 1260 ttgctggaagagcgtgactatattgagcgttggagc 1294  
 Db 1073 CTGCAGCGAGATGCGACGATATGCGCATGGGATG 1107

RESULT 7  
 AX090448 1518 bp DNA linear PAT 21-MAR-2001

LOCUS AX090448  
 DEFINITION Sequence 37 from Patent WO0116325.  
 ACCESSION AX090448  
 VERSION AX090448.1 GI:13444286

KEYWORDS  
 SOURCE  
 ORGANISM  
 bread wheat.  
 Triticum aestivum

REFERENCE  
 AUTHORS Butler,K.H., Danilevskaia,O., Miao,G.H., Morgante,M., Sakai,H.,  
 1 (bases 1 to 1518)  
 Simonon,C.R. and Weng,Z.  
 Plant reproduction proteins  
 Patent: WO 0116325-A 37 08-MAR-2001;  
 E.I. DU PONT DE NEMOURS AND COMPANY (US) ; Pioneer Hi-Bred  
 International, Inc. (US)  
 TITLE  
 JOURNAL

FEATURES  
 source  
 1..1518  
 Location/Qualifiers

/organism="Triticum aestivum"  
BASE COUNT 393 a 337 c 403 g 384 t 1 others  
ORIGIN

Query Match 31.9% Score 498.2; DB 6; Length 1518;  
Best Local Similarity 65.3%; Pred. No. 2.5e-120;  
Matches 731; Conservative 0; Mismatches 388; Indels 0; Gaps 0;

Oy 214 ttagggaacgaatcaatagttggtcttctgacatcgaataagaacgtacaaagt 273  
Db 115 ttagggtcgagagcgccgggtgggtcgctggcgccacgccggaagccgagatcaagctc 174  
Oy 274 acgaatagatcaggaagaagaacacctgtatgtctgtgttttcaactcctgac 333  
Db 175 tgcacacgaacacacgagggcggccgctctacgagccatgggttcaattcatcgac 234  
Oy 334 gctgtctctcgatgctgtctacgctgtgtgaatacsgatcctctgtacaaattgt 393  
Db 235 gccctctactacgagcttctccacccgctcgccggaatgctgacgacgacgacgacg 294  
Oy 394 ctggaagatgtgcatatcaacatgcatcctatgtctgtatgaagaagaagatcg 453  
Db 295 ctcccgacagagatcttgctgtctgacgaacatgacatgacgacgacgacgacgacg 354  
Oy 454 tttaacacgtaagtgtggcgctgtgcgttaataaggaaccataatgtgcgctgagga 513  
Db 355 ttctacacactgagctggcgctgtgtgacgtgacgacgacacacgctgacgacgacg 414  
Oy 514 gtaaaagtataatccgaatcatgtacgtcaacagatgaacatcatatgaagctgtg 573  
Db 415 accaattgcgctattcggtctcatcactgtccacacgagaaagtgttttagagtttctt 474  
Oy 574 ggtcatggaatcagtgtaacgaacaaacacacacacacacacacacacacacacac 633  
Db 475 gcccatggtgattcatataatagatgaacacacacacacacacacacacacacacac 534  
Oy 634 actgctgacgaagatgaatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 693  
Db 535 tctgcagacgaagacgagctgtgtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 594  
Oy 694 atattgtgagagctggaagctgcatgaatgttaagttgagttgatttccgtt 753  
Db 595 atttttgcctggagagagacgacgtatgaatgagttgagttgacacacacacacac 654  
Oy 754 gatattacacgcttgcagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 813  
Db 655 gatattcaccgaattgtgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 714  
Oy 814 gagtttgaagctgacgtgcgaagaatcacaatcgaatcgaatcgaatcgaatcgaat 873  
Db 715 gaattttggccctgacgt 774  
Oy 874 aaaaaattgtccatcctcgttattacagcttccatcacaataatattagattt 933  
Db 775 acgaattttgtcaatttccgtctcatgacttccgtgttcatcttcatctgtgtgtgt 834  
Oy 934 aacgctgt 993  
Db 835 actagtggt 894  
Oy 994 gaacacacacgtgaagaagaatcctcctgctgaggaagcttcagatgtctatagaatc 1053  
Db 895 gagcccaaaaataaagagagaggtcccgccgagggtagcatgattgttcttgaagaatg 954  
Oy 1054 ccggttccaagtgtgtatattgtgttttcaagtttctgtgtgtgtgtgtgtgtgtgt 1113  
Db 955 cctgtgcccgt 1014  
Oy 1114 gttgtgataagtaacgaagaagaagtttactgtgtgtgtgtgtgtgtgtgtgtgtgt 1173  
Db 1015 tttagcaatttagcgaacgacgaagcaaaatctatgtgtgtgtgtgtgtgtgtgtgtgt 1074

Oy 1174 gttttattacaagaattatcacacaatcaaatcaaatcgttaatcaggaacagccatg 1233  
Db 1075 gtgcataattaccgctgtagtccatcaatcaatcaatcaatcaatcaatcaatcaat 1134  
Oy 1234 tctgtcattggaagacagatcttctgtctgtcgtggaagacgagataatgagcgtggac 1293  
Db 1135 tgcgtttgatggaagacagatcttctgtctgtcgtggaagacgagataatgagcgtgg 1194  
Oy 1294 gtgattaccgaatgagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1332  
Db 1195 gaagtgaacatcaacgtcgaataaaattgagacgacgacgacgacgacgacgacg 1233

RESULT 8  
AX090412  
LOCUS AX090412 1643 bp DNA linear PAT 21-MAR-2001  
DEFINITION Sequence 1 from Patent WO0116325.  
ACCESSION AX090412  
VERSION AX090412.1 GI:13444268  
KEYWORDS  
SOURCE  
ORGANISM  
Zea mays.  
Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
1 (bases 1 to 1643)  
Butler, K.H., Danilevskaia, O., Miao, G.H., Morgante, M., Sakai, H.,  
Simmons, C.R. and Weng, Z.  
Plant reproduction proteins  
Patent: WO 0116325-A 1 08-MAR-2001;  
E.I. DU PONT DE NEMOURS AND COMPANY (US) ; Pioneer Hi-Bred  
International, Inc. (US)  
Location/Qualifiers  
1..1643  
/organism="Zea mays"  
/db\_xref="taxon:4577"

BASE COUNT 429 a 345 c 456 g 413 t  
ORIGIN

Query Match 31.3% Score 489.4; DB 6; Length 1643;  
Best Local Similarity 66.4%; Pred. No. 5.2e-118;  
Matches 719; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

Oy 215 taagggaacgaatcaatagttggtcttctgacatcgaataagaacgttacaagttga 274  
Db 221 tgcggtgcagagggcgaggggtgcctgcgcacgacggaagggagatcaaacgctt 280  
Oy 275 cgaatagatcagaagaagaagaacaccttgatgctgtgtgtgtgtgtgtgtgtgtgtgt 334  
Db 281 gcggcgaagacacactgtgaggggaacggcgcctatattcttctggttcaacttcatgag 340  
Oy 335 ctgcgttctcgtatgctgtcgttaccgctgtgtgtgaatcgaatcgaatcgaatcgaat 394  
Db 341 cgcgctactacacagcttcttccacccgctgcgacgaacccgctgacaaactttacccgtgcc 400  
Oy 395 tcggaagatgtgccatatacgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 454  
Db 401 ttgagaattgtagttgttcgcttcttcttcttcttcttcttcttcttcttcttcttctt 460  
Oy 455 tttaacagtgtaagtgtggcgctgtgc---gttaatggaacccatattgttcgctgtgag 511  
Db 461 tctattctcttaagctgtggcgtgtgcacacatgtgattggtccaccacgctgtgtgtgtg 520  
Oy 512 ggttaaaaggtataatccgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 571  
Db 521 gaagcaattgggacatcttccgggtcatcaattgtgtctacagaaagttacctgaagacgtttg 580  
Oy 572 tgggtcattggaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 631  
Db 581 ttggccatggcgcgtcaataaattgagattgaagaaactcaaccggttggaagccttgcctatca 640

QY	632	ttatgtctagaagaagaaatctgttcgtttctgttggaatgttgaacgtgatatctt	691
QY	632	ttatgtctagaagaagaaatctgttcgtttctgttggaatgttgaacgtgatatctt	691
Db	641	ttttctgaagcaagagatgaatctgtttgagcttngtgaaatgttccatctacggagatctgatact	700
QY	692	tgaattctgcgagcgagagatcatgcgtatagaagttcaagtgttgatttcaatcgt	751
Db	701	tgaattttgtcggagcgaggagatcatgcgtatagaagttcaagtgttgatttcaatcgt	760
QY	752	ctgaattaccgcgtctgttagttgtgtgtatgagacaccactatataatgtgtcaatga	811
Db	761	gtgatatttgaaacgtttttgcaagttgtgtgacatgcgaacacactgtgaaaaatctgtgcaatga	820
QY	812	aagagttttgagacgtacgtgagagagatcatcatcagatgacgtatgacatcaaatctcc	871
Db	821	aagattttttgtgctatgattgttgcacaaatcatattttatgatgactgaccttccatcaaaagtttc	880
QY	872	ccacaagaattgttccaaattccctgtlattacagctttccatcatcacaatathatgtagatt	931
Db	881	caacaaatattgttcacagtttccagatcttgatgtgctgcagatgcacgtcctaaatgttgatt	940
QY	932	gtaacgctgtgtgtgtatattatcctctcaaaagatgtgtgacaaagagatcctgtgtc	991
Db	941	gtacaaagatgacctgtgtgacttcaatcctaatcaaaagagttgacaaatgaaattgtgctttt	1000
QY	992	gggaaccccaactgaagaagaatctcctctgcgagggagcttcagatgtctatlaaagat	1051
Db	1001	gggaaccccaactgaagaagaatctcctctgcgagggagcttcagatgtctatlaaagat	1060
QY	1052	accggtgtccaaatgtgtatatttgattcatcaagtttctgtgacctcatlaaagtt	1111
Db	1061	atcctgtgtcccaaatgtgatactttgtttatcaaatttttcattgtatattttcacttcaatc	1120
QY	1112	ctgttcgataagtaatcaaggaagaagttatgtctgtggaatttgaagaattgtccctc	1171
Db	1121	agttgcgcatagcgaacgctgaagcgaacaaatctcagtgctgggaatgtacagtcacacccctc	1180
QY	1172	ctgttttgattacaagaattatcaccaatcaatcaaatctgtatctgatactggagaagaagca	1231
Db	1181	ctgtcctcatcttccctgcgctgataatcatcagcaatctgtaattcggcgaatgaacaaatctgac	1240
QY	1232	tgtctgtcagatgaagacagatctctgtctgttcgagagcggagctatgtgcgctgag	1291
Db	1241	tttcttcttcgattgaaagcacaaatctttggagctgtgtgaagacggcgcacaaatctgtgcgctgag	1300
QY	1292	acg 1294	
Db	1301	atg 1303	
RESULT	9		
AX090440		1686 bp	DNA
LOCUS	AX090440		linear
DEFINITION	Sequence 29 from Patent WO0116325.		
ACCESSION	AX090440		
VERSION	AX090440.1	GI:13444282	
KEYWORDS			
SOURCE			
ORGANISM			
	Zea mays.		
	Zea mays		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC		
	clade; Panicoideae; Andropogoneae; Zea.		
	1 (bases 1 to 1686)		
REFERENCE			
AUTHORS	Butler,K.H., Danilewskaya,O., Miao,G.H., Morgante,M., Sakai,H.,		
	Simmons,C.R. and Weng,Z.		
	Plant reproduction proteins		
	Patent: WO 0116325-A 29 08-MAR-2001;		
	E.I. DU PONT DE NEMOURS AND COMPANY (US) ; Pioneer Hi-Bred		
	International, Inc. (US)		
	Location/Qualifiers		
	1..1686		
FEATURES			
source			
	/organism="Zea mays"		
	/db_xref="taxon:4577"		
BASE COUNT	425 a	374 c	468 g 419 t

Query Match	31.3%	Score 489.4	DB 6	Length 1686
Best Local Similarity	65.4%	Pred. No. 5.2e-118		
Matches 719	Conservative	0	Mismatches 361	Indels 3
				Gaps 1
QY	215	taggaagcagatcacaatgttggtcttcttgacccctcgaataagaatcgtacaaagtga	274	
DB	422	TCGGGTGGCAGGCGCGCGAGGGGTGCGTCGTGCCACGGGAGAGCGGAGTACAAAGCCCT	481	
QY	275	cgaatcaggaatcaagaagggaaacaccttgaatgctgtgtgttlltcaactccttgatg	334	
DB	482	CGGCGAAGCACACTGAGGGGAGCGCCCGCTATATGCTATCGGTTCAACCTTCAGAGAG	541	
QY	335	cctcgtttcttgatgtcttctgttaccgctgtgagaaatcggatctctcgtacaaatctc	394	
DB	542	CCCGCTACTACACAGCTCTTCGCCCGCTGGCGGCAACCGCTGACAACTTACCGCTGCC	601	
QY	395	tggagatgtgtccatcagcatctgcaatctcctatgctatgaataagaagaagatcgt	454	
DB	602	TTGAGATGTGATGTTTCGCTCTTCTTACAAAGCTTACGTTGATGAGATTAAGATGATGCT	661	
QY	455	tttacacggtaaagtbtgggcgtgtggc--gtlaatggaaacccatatgttgcgctgtgag	511	
DB	662	TCTATCTCTTAAGTGGGCTCGGACCATGTGATGCTGCACCACTGCTGAGGAGAG	721	
QY	512	gagtaaaaggtabaatccgagtcattgtacgtcaacaggtgaacagatctcaataagatctg	571	
DB	722	GAAGCAATGGGATCATTCGGGTCATCAATTTGCTACAAAGAAAGTTAGCTTAAGAGCTTGG	781	
QY	572	tgggtcatgtgagatttcgaagcgaatcaagaacacacaccttaaacctcaactcgttga	631	
DB	782	TTGGCCATGGCGCTCAATTAATGAATTAAGAACTTAACCGTTGAAGCCTTGCTCATCA	841	
QY	632	ttactgtcagaaggaatgaatcgttctgtgtgtgtgaatgttgaacttggatgatalatt	691	
DB	842	TTTCTGCAAGCAAGATGATGATGTTAGGGGTATGGATGGATGCCATACAGGATCTGATCT	901	
QY	692	tgaattctgtcgtgagcttggaggtcactcgtctatagaagtcttaagtgtgatttccatcgt	751	
DB	902	TGATATTGCTGGAGCTGGAGGTCAATCGCATGCAATGATGATGATGATGATGATGATGATG	961	
QY	752	ctgatatctccgcttctgtcagttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	811	
DB	962	GTAATATTGAACTTTTGGCAAGTTTGGCATTTGGCATGTGACACAACTGTGAAATGTGTCATGA	1021	
QY	812	aagagtttggagctacgtcgcgaagatctatcactcagatgtgacatgtatgacatcaaaatcc	871	
DB	1022	AAGATTTTGGCTATATGTTGACAAATCATATTCATGAGTGCATCTCCATCAAGTTTC	1081	
QY	872	cccaaaaatttgcacaatccctgtlatcttaacagcttccatctcaataaatatgtatgatt	931	
DB	1082	CACCAAAATATGTCACAGTTTCCAGTCTTGTGATTCCTGACATCACTCAACTATGTTGATT	1141	
QY	932	gtaacggt	991	
DB	1142	GTAAAGATGAGTGTGGTGTGACTTCACTTATCAAGATGTGTTGACATGAATATGTGCTTT	1201	
QY	992	ggagacacacagaaagaaatctctcgtcgagagagatctcagatgtctatlaaat	1051	
DB	1202	GGGAAACCGAAGACAAAGAACAGAGTCTGGGGGAGGAGACATCGATATCTTTCAGAGT	1261	
QY	1052	accggttccaatgtgtgataatttggttatcaagtttcttgtgacctccatlttaagt	1111	
DB	1262	ATCTGTGCCAGATGTGACATTTGTTTATCAATTTTATGATGATTTTCACTTCATATC	1321	
QY	1112	ctgttgcgataagtaatacaggaagaaagtttatgtctcgtgatttgaagaagtctccctc	1171	
DB	1322	AGTTGGGATGTGGCAACCGTCAAGGCAAAATCTATGCTGGGAAGTACAGTCCAGCCCTC	1381	
QY	1172	ctgttctgattacaagaattatcacacaatcaatcaagaatcgtatcaaggcaaacagca	1231	

Db 1382 CTGCTCATTTGCTGGCTGTATTAATCAGCAGTAAATGCCGATAGACAAACTGCAG 1441  
QY 1232 tgtctgtcgatggaagcaagatcttgcctgtcgcgagaagcgactataatgctgtgg 1291  
Db 1442 TGTCTTGTGATGGAACACAAATCTTGAGCTGTGGAAAGACGACCATCTGGCGTGGG 1501  
QY 1292 acg 1294  
Db 1502 ATG 1504

RESULT 10  
AY061965 1711 bp mRNA linear PLN 01-JAN-2002  
LOCUS zea mays fertilization-independent endosperm protein (FIE2) mRNA,  
DEFINITION complete cds.  
VERSION AY061965  
KEYWORDS  
SOURCE zea mays.  
ORGANISM zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoidae; Andropogoneae; Zea.  
1 (bases 1 to 1711)  
Phillips,R.L., Kaeppler,H.F. and Kaeppler,S.M.  
Sequence Relationships, Conserved Domains, and Expression Patterns  
for Zea mays Homologs of the Drosophila Polycomb Group Genes E(z),  
esc, and E(Pc)  
Plant Physiol. (2002) In press  
2 (bases 1 to 1711)  
Danilevskaya,O.N. and Springer,N.M.  
Direct Submission  
Submitted (07-NOV-2001) Agronomic Traits, Pioneer Hi-Bred Intl.,  
Inc, 7250 NW 62nd Ave, Johnston, IA 50131-0552, USA

FEATURES  
source  
1..1711  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
/chromosome="10"  
/map="bin 10.3"  
/note="genotype: B73"  
1..1711  
/gene="FIE2"  
/gene="FIE2"  
394..1533  
/note="WD polycomb group protein; similar to Arabidopsis  
thaliana fertilization-independent endosperm protein FIE"  
/codon\_start=1  
/product="fertilization-independent endosperm protein"  
/protein\_id="AAL35974.1"  
/db\_xref="GI:18032006"  
/translation="MAKLGPGGLGCEAEGLSVSRREKPKCGKHTGKRLPLAIG  
FNFMDARYDEVATYVGNRTYRLEGLGALLQAYDEDDKDEFYTLNARHDVDC  
SPILVAAGSNGLIIVINCATEKLAKSPFGHDSINEITQPLKSLIISAKDSDEVR  
KNVHTGICILIFPAGAGHREVLSDYRHSDEIRPAGSGMDYFAVIMSKKEFWLVD  
SYSDTDFSKFPTKYVQDPVLAIAVHNSVYCTRWLQGLFSLKSYDNIWMEPTKE  
QSPGSDIDLOKYPVEPCDIFPKFSCDFHNOALIGNREKLIYVWVQSSPVLIA  
RLYNQOCSPIRQTAIVSFDSITLIGAGEDGIMRWDEVDPHSSRN"

BASE COUNT 450 a 374 c 468 g 419 t  
ORIGIN

Query Match 31.3% Score 489.4; DB 8; Length 1711;  
Best Local Similarity 66.4%; Pred. No. 5,2e-118;  
Matches 719; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

QY 215 taggaaagagatcaatagtgtgactccatcagaataagaatgatacaagtga 274  
Db 422 TCGGGGTGCGAGCGCGGAGGGGTGCTGTGCCAGCGGAGCGGAGTACAAAGCCT 481  
QY 275 cgaatagaattcagaaggaagaaaccttgatgctgtgtttcaactccctgatg 334

Db 482 GCGGACAGACACTGAGGGGAAGCCGCCGTATATCTATCGCGCTCAACTTCATGAGAG 541  
QY 335 ctgcgttctcgatgctcttcgttaccgcgtgtggaatcggatctactgtaaatgtc 394  
Db 542 CCGGCTACTACGACGCTTCCGACCGTGGCGGCAACCGCGTGACAACTTACCGCTGCC 601  
QY 395 tcggagaatggtccatatacaagcatcgaatcctatgctgataagaataagaagtcgt 454  
Db 602 TTGAGAAATGTAAGTTTCGCTTCTTCAAGCTTACCTGATGAGATGAAGATGATGCT 661  
QY 455 ttacacggttaagtgtggcgtgtgc---gttaatggaacccaatgtgtgcgcggag 511  
Db 662 TGTATCTCTAAGCTGCGCTGCGTGCATGTTATGATGCTACACACGCTGTTGGCAGAG 721  
QY 512 gagtaaaagatatacgcagatcattgacgtcaacagtgaaacgattcaagaagtcgt 571  
Db 722 GAAGCAATGGATCATTTCCGGTCAATGTTGCTACAGAAAAGTAAAGCTTAAGAGCTTGG 781  
QY 572 tggatcatggaattcagtgaaacgaatacaagacaacaccttaaacctcaactgtga 631  
Db 782 TTGGCCATGGCGACTCAATTAATAGATTAAGAACTCAACCGTTGAAGCCTTGCCTATCA 841  
QY 632 ttactgtcagaagatgatactgttctgtgtgtggaatgtggaactggatgatatatt 691  
Db 842 TTTCTGCAAGCAAGATGATGATCTGTAGGCTATGGAATGCAATKACAGGATCTGATCT 901  
QY 692 tgatatgtcgtgaagctggaagtcacgcgtatgaagttcaagtgtgattcaccgt 751  
Db 902 TGATATTTCTGGAGCTGGAGGTCATCGCAATGAATGATTAAGTGTGACTTCCATCTTA 961  
QY 752 cgtatatccagcttctgt 811  
Db 962 GTGATATTAAGCTTTTGGAAAGTTGGCATGACAACTGGAATAATCTGTGATATGA 1021  
QY 812 aagaagtgtggaagctgcgcgaagatcattcaacagagctgatagtcatcaaaattcc 871  
Db 1022 AAGAAATTTGGCTATATGTTGACAAATCATATTCAATGACATGACCTTCCATCAAAATTTC 1081  
QY 872 ccacaanaattgtccaattccctgtatattacagcttcattcaataaattatgtgatt 931  
Db 1082 CAACAATAATATGTCAGCTTCCAGCTTGATTCAGTCACTGATACACTTAATCTGATTT 1141  
QY 932 gtaaccgt 991  
Db 1142 GTAACAAGATGGCTTGTGTGATCTCATCTATCAAAAGAGTGTGCAATGAATTTGCTCTT 1201  
QY 992 ggaaccacaacatgaaagaaattcctcctggaaggaagcttcaagatgttctaatgaat 1051  
Db 1202 GGGAAACCGAAGACAAAGAACAGAGTCTGGGAGGGAAGCATCGATATCTTCAGAAAT 1261  
QY 1052 acccggttccaattgtgtatattgtttatcaagttcttctgtgacccatcaattagt 1111  
Db 1262 ATCCGTCTCCAGAAATGTGACATTTGGTTATCAAAATTTTCATGATGATTTCACTTCAATC 1321  
QY 1112 cgttgagatggtatcaagaaagaaagttatgtctgggaattgaaaagttgcctc 1171  
Db 1322 AGTTGGCGATAGGCAACCTGTGAAGGCAAAATCTACCTGTGGGAGTACAGTCCAGCCCTC 1381  
QY 1172 cgttttggattcaaaagttatcaacacatcaatcaagaactctgtaacaggaagagca 1231  
Db 1382 CTGCTCTCATTTGCTGGCTGTATTAATCAGCAGTGTAAATCGCCGATTAACAAACTGCAG 1441  
QY 1232 tgtctgtcgatggaagcaagatcttgcctgtcgcgagaagcgactataatgctgtgg 1291  
Db 1442 TGTCTTGTGATGGAACACAAATCTTGAGCTGTGGAAAGACGACCATCTGGCGTGGG 1501  
QY 1292 acg 1294  
Db 1502 ATG 1504

RESULT 11







OY	454	ttttaaccggtlaaglttgggc---	glgttgacgtttaatlyggaaaccataatgttcgctgtga	510
Db	572	TTTACACTCTAACCTGGACCATGCATCAAGTTGATAGCTCACCGCGTGTTGGTGGCCGA	631	
OY	511	ggagttaaaggatataatccgaatcatlgaagtcacaacgtigaacgattcaagaagtcct	570	
Db	632	GGAAGCATTCGGATCATTCCGGGTCAATAATTTGCTTCCGAAAGTTAGATTAAAGACTTA	691	
OY	571	gtgggtcattgtagatatgaagtaacgaataatccagacacaaccttaaacctccaactgtg	630	
Db	692	GTTGGCCATGTGTGTTTCANTACATGAGATAGAAGCATCATGCTGCCTCGAAGCCATCAGTCATC	751	
OY	631	attaactgtacgaaggatgaatcctgttcgttttgtgaaagtltgaactcygatatyat	690	
Db	752	ATTCTGCCACACAAGATGATATATTAGCGTATAGGAATGCTCATCTGGGATTTGGATC	811	
OY	691	ttgatatattcgctgtgaagctgtgaagtcacatgcgtatgaagtlctaagltfvggaattlccacg	750	
Db	812	TTAGCTTTTGGCAGGGGCGTGGAGGCCATCGACATGATGTTGAGATGTTCACCTCCACCT	871	
OY	751	tctgatlattacocgttctgtcaatgtgtgatatgtgaaaccaataataaatatgylcaatg	810	
Db	872	AOCGAGTGTGGGAMTTTTTGCAGATGTGTGGCATTTGCANACTCTGTGAAAATTTGGTCAATG	931	
OY	811	aagaagatttgagacgtacgtcgaagaagcattccaatgyaactgtgaatgacatcaatlc	870	
Db	932	AAAGATATTTTGGATATATGTTGAAAAATCATATTCATGATGAGCCATCCATCAAAGTTT	991	
OY	871	cccacaaaattgtccaattccctgttatattacagcttccatccaataaatatagiat	930	
Db	992	CCAAGCAGGAATATTCAGTTTCCGGCTTTGACGTGTCAGTACACTCTGACTATATGTTGAT	1051	
OY	931	tgtaacgttggtttggatatttacctccaagaagtgtygcaaacgagatcctgtg	990	
Db	1052	TTTACAGATGTGCTTGtGtACTTCACTCTATCAAGAAGATGAAGAATGACAGTTTGGCTT	1111	
OY	991	tgggaaccacactgaagaagaatctcctgtgcgaaggagcttcagatgltctaltaaga	1050	
Db	1112	TGGGAACCAAAACNAGACAAAGCGCTGGCGTGGGAGGAGTGTGATGTTCTTCAGAAAG	1171	
OY	1051	taccgggttccaaatgttgyataatttggatttcaaaagtttcttggactccatttaagt	1110	
Db	1172	TACCCGCTGCCAAAGTGTTCAATTATGAGTTTGAATAATTTCAATGTGATTTTACTCCAAC	1231	
OY	1111	tctgtgtgcatgtgtaatacgaagaagaaggtttagtctgtgagatttgaagaagtgcct	1170	
Db	1232	CAGATGGCATATAGGCCAACAAATAAAGCCAGATCTATGCTGTGGGAAGTGCAGTCCAGCCG	1291	
OY	1171	cctgttctgattcaaaagtatcacaaatcaatcaaatcaaatcgttaatacgaagcaaacg	1230	
Db	1292	CCCCGTTTAAATTGACCGGCTGTGCMAACCAAGAAATGGAATGCCATTAAGCAGACCCGA	1351	
OY	1231	atgtctgtcattgaaagacgaatcttctgtctgtcgcgagagcggacataatgqcgctgg	1290	
Db	1352	GTTCTATTCGACGAGCAAGATCTTGGAGCGCCGACGAGCGGCGATCTCTGCGGGTGG	1411	
OY	1291	gacg 1294		
Db	1412	GACG 1415		
RESULT	15			
LOCUS	AY061964	1787 bp	mRNA	linear
DEFINITION	Zea mays fertilization-independent endosperm protein (FIE1) mRNA,			
complete cds.				
ACCESSION	AY061964			
VERSION	AY061964.1	GI:18032003		
KEYWORDS	.			
SOURCE	Zea mays.			
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			











RESULT 3  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:

1 APPLICANT: DORNER, F.  
 2 APPLICANT: SCHEIFLINGER, F.  
 3 APPLICANT: FALKNER, F. G.  
 4 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 5 NUMBER OF SEQUENCES: 52  
 6  
 7 CORRESPONDENCE ADDRESS:  
 8 ADDRESSEE: Foley & Lardner  
 9 STREET: 1800 Diagonal Road, Suite 500  
 10 CITY: Alexandria  
 11 STATE: VA  
 12  
 13 COUNTRY: USA  
 14 ZIP: 22313-0299  
 15  
 16 COMPUTER READABLE FORM:  
 17 MEDIUM TYPE: Floppy disk  
 18 COMPUTER: IBM PC compatible  
 19 OPERATING SYSTEM: PC-DOS/MS-DOS  
 20 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 21  
 22 CURRENT APPLICATION DATA:  
 23 APPLICATION NUMBER: US/08/232,463  
 24 FILING DATE:  
 25 CLASSIFICATION: 435  
 26 PRIOR APPLICATION DATA:  
 27 APPLICATION NUMBER: US/07/935,313  
 28 FILING DATE:  
 29 APPLICATION NUMBER: EP 91 114 300.6  
 30 FILING DATE: 26-AUG-1991  
 31 ATTORNEY/AGENT INFORMATION:  
 32 NAME: BENT, Stephen A.  
 33 REGISTRATION NUMBER: 29,768  
 34 REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
 35 TELECOMMUNICATION INFORMATION:  
 36 TELEPHONE: (703)836-9300  
 37 TELEFAX: (703)683-4109  
 38  
 39 TELEX: 899149  
 40  
 41 INFORMATION FOR SEQ. ID NO.: 14:  
 42 SEQUENCE CHARACTERISTICS:  
 43 LENGTH: 7218 base pairs  
 44 TYPE: nucleic acid  
 45 STRANDEDNESS: single  
 46 TOPOLOGY: linear  
 47  
 48 IMMEDIATE SOURCE:  
 49 CLONE: pTZ19pt-Flis  
 50  
 51 US-08-232-463-14

```
Query Match      2.5%; Score 38.4; DB 1, Length 7218;
Best Local Similarity 1.9%; Pred. No. 0.14;
Matches 6; Conservative 180; Mismatches 126; Indels 0; Gaps 0;
```

RESULT 4  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Db 12077 TATCAATGAGTGTATAATTTGCTCTTCTGCGATTGGGTTACCTCACTCAGGATGAT 12136

Matches 87; Conservativ

capre

Oy 907 tcaatcaacaataatgtagatgaaacgttgittggatatttaactccaag 966  
 Db 43310 tactctgtaggaaatgaagagatagtaataagaaattcgggttcccttttaaaaa 43365  
 Oy 967 agtgcggacaagagatccgtctggggacacaactgaaagaaattccctgcgag 1026  
 Db 43370 aattcagcttgcatattccaaaggccaattcatgatttaaaaaaatttcccttgct 43425  
 Oy 1027 ggaagcttcagatgtctcatlaagaaccggcttcaatgtagatatggg 1078  
 Db 43430 tgcagtgacaatgatttttttagacatcctgctgtttagcaagttaattttgtg 43481

RESULT 8  
US-08-972-927-4  
; Sequence 4, Application US/08972927  
; Date of Issuance 01/05/2008

```

: PATENT NO. 50168290
:
: GENERAL INFORMATION:
: APPLICANT: Rea, Phillip A
: APPLICANT: Lu, Yu-ping
: APPLICANT: Li, Ze-sheng
: TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
: TITLE OF INVENTION: PLANTS
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
: STREET: One Commerce Square, 2005 Market Street, 22nd
: STREET: Floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: US
: ZIP: 19103-7086
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/972,927
: FILING DATE: 18-NOV-1997
: CLASSIFICATION: 800
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/031,040
: FILING DATE: 18-NOV-1996
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/061,328
: FILING DATE: 08-OCT-1997
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Doyle Leary Ph.D., Kathryn
: REGISTRATION NUMBER: 36,317
: REFERENCE/DOCKET NUMBER: 9596-1202
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-965-1284
: TELEFAX: 215-567-2991
:
: TELEX: 831-494
:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5175 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: MOLECULE TYPE: cDNA
:
: US-08-972-927-4

```

	Query Match	2.2%	Score 34.8	DB 4	length 5175
	Best Local Similarity	51.3%	Pred No. 1.5		
	Matches	81	Conservative	0	Mismatches 77; Indels 0; Gaps
QY	599	tcagacacacaccttaaacctcaactgtgtactgtcgtgcgaagatgaatctgttc	658		
DB	329	tcaggctcttAAAGGTTTACCTATTTTCTGGCCCTTTTGGCTGCTGTTATGTACTGCGGAC	388		

Oy	659	gttgtagaacgcttgaacacggagatagtatttcgatattgcgcggagcgaggatc	718
Db	389	ctttctttaagtttgtctcatagagactctctgttttgcaatttggatgacagctgcgcttcc	448
Oy	719	gctatgaagttctaagtggtgatttcaatccgtctgat	756
Dd	449	cctatgagcgccgttatgttggtccttgagcgctttttgct	486

RESULT 9  
US-08-945-056-4  
; Sequence 4, Application US/08945056

1 Patent No. 6077994  
 2 GENERAL INFORMATION:  
 3 APPLICANT: Coupland, George M.  
 4 APPLICANT: Puterill, Joanna J.  
 5 TITLE OF INVENTION: Genetic control of flowering  
 6 NUMBER OF SEQUENCES: 31  
 7 CORRESPONDENCE ADDRESS:  
 8 ADDRESSEE: Nixon & Vanderyhe PC  
 9 STREET: 8th Floor, 1100 No. 6077994th Glebe Road  
 10 CITY: Arlington  
 11 STATE: Virginia  
 12 COUNTRY: United States of America  
 13 COMPUTER READABLE FORM:  
 14 MEDIUM TYPE: Floppy disk  
 15 COMPUTER: IBM PC compatible  
 16 OPERATING SYSTEM: PC-DOS/MS-DOS  
 17 SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)  
 18 CURRENT APPLICATION DATA:  
 19 APPLICATION NUMBER: US/08/945,056  
 20 FILING DATE: 20-OCT-1997  
 21 CLASSIFICATION: 800  
 22 PRIOR APPLICATION DATA:  
 23 APPLICATION NUMBER: PCT/GB95/02561  
 24 FILING DATE: 01-NOV-1995  
 25 PRIOR APPLICATION DATA:  
 26 APPLICATION NUMBER: GB 9422083.7  
 27 FILING DATE: 02-NOV-1994  
 28 ATTORNEY/AGENT INFORMATION:  
 29 NAME: Ms Mary J Wilson  
 30 REGISTRATION NUMBER: 32,955  
 31 REFERENCE/DOCKET NUMBER: 620-17  
 32 TELECOMMUNICATION INFORMATION:  
 33 TELEPHONE: (703) 816-4000  
 34 INFORMATION FOR SEQ ID NO: 4:  
 35 SEQUENCE CHARACTERISTICS:  
 36 LENGTH: 4201 base pairs  
 37 TYPE: nucleic acid  
 38 STRANDEDNESS: single  
 39 TOPOLOGY: linear  
 40 MOLECULE TYPE: DNA (genomic)  
 41 ORIGINAL SOURCE:  
 42 ORGANISM: Arabidopsis thaliana  
 43 STRAIN: Landsberg erecta  
 44 POSITION IN GENOME:  
 45 MAP POSITION: chromosome 5  
 46 US-08-945-056-4

Query Match	2.2%	Score 34.6	DB 3	length 4201
Best Local Similarity	46.5%	Pred. No. 1.6	Mismatches 129	Indels 0
Matches 112	Conservative	0	Gaps 0	
QY	138	agactgtgagattagattagctgtgcagagtcacacagagagagattccaatcgc	197	
Db	3912	AGAGATAGACCCGGGGTCAATGCGCGTTCGCAAGAAGAAATCGACCGGAGAGCA	3971	
QY	198	aatgtcgaagataaccttagggacgagtcacaaatagtttgctttgactcatcaataa	257	
Db	3972	AGGGTTCAACACGAGATGCTAATGTACAAACAGAGATATGGATTGTCTTCATTGATA	4031	
QY	256	gaatatcgtacaaagtgcagcaatagattcaggaaaggaaacctttgatgctgttgt	317	

Db 4032 CTCCTGTGCAAAAAGAAACATGATTCAGACGTGTAATTAATTACTTTAGCTTGATTA 4091  
OY 318 ttcaacttccttgatgtgttcttcgaigtcttcgttaccgctgtgtgaatcgat 377  
Db 4092 TGTTAGGTTGGTGAATTTAGCTTCAGAGATTAATTACTTGTGTGCAAAAGGTT 4151  
OY 378 t 378  
Db 4152 T 4152

## RESULT 10

US-09-308-003-4  
Sequence 4, Application US/09308003  
Patent No. 6326170  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K. R.  
APPLICANT: Lonetto, Michael A.  
APPLICANT: Warren, Patrick V.  
TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,  
TITLE OF INVENTION: POLYPEPTIDES AND THEIR USES  
FILE REFERENCE: GM10093  
CURRENT FILING DATE: 1999-05-10  
EARLIER APPLICATION NUMBER: 60/058,710  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 990  
TYPE: DNA  
ORGANISM: Staphylococcus aureus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(987)  
US-09-308-003-4

Query Match 2.2%; Score 34.4; DB 4; Length 990;  
Best Local Similarity 57.4%; Pred. No. 0.84;  
Matches 62; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
OY 432 tgaatgaagaagaagagtcgtttacacgtaagttggcgctgtgtgaatggaatgaa 491  
Db 18 tgaatgaagaagaagagcgctgtctctatgcccacagctgtgttgaatggaatgaa 77  
OY 492 cccatatgttcggtcgtgaggaagtaaggtatataccgagtcattga 539  
Db 78 ccaaatgttaagcagaagaataaataaagtaacgaagtcattaa 125

## RESULT 11

US-08-271-562-2/c  
Sequence 2, Application US/08271562  
Patent No. 5573940  
GENERAL INFORMATION:  
APPLICANT: Sims, Peter J.  
APPLICANT: Bothwell, Alfred L.M.  
TITLE OF INVENTION: Genetic Inhibition of Complement  
TITLE OF INVENTION: Mediated Inflammatory Response  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick & Cody  
STREET: 1100 Peachtree Street, Suite 2800  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30309-4530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/271,562  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/729926  
FILING DATE: 15-JUL-1991  
APPLICATION NUMBER: US 07/365199  
FILING DATE: 04-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: OMR# 112CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 815-6508  
TELEFAX: (404) 815-6555  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 315 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Blood  
CELL TYPE: Erythrocyte  
US-08-271-562-2

Query Match 2.1%; Score 33.4; DB 1; Length 315;  
Best Local Similarity 51.7%; Pred. No. 0.93;  
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
OY 669 tgttgaacgtgagatagttatgtatctgtgagcgtgagagtcatactgataagt 728  
Db 196 tcttgacgacgtagttagcttcttccctcaacggggttgacgcttgtaaat 137  
OY 729 tctaatgtgatttcatcgcgttcgatattacccgcttgcagtgtgtatgagacac 788  
Db 136 tgcagtcctcaaaccttcacacactgttatcacactgttaaccacgcttggatgagac 77  
OY 789 cactataaataatggtcaatgaaga 815  
Db 76 acgcattcaaatcagatgaacattga 50

## RESULT 12

US-08-087-007-4/c  
Sequence 4, Application US/08087007  
Patent No. 5705732  
Patent No. 5705732 5684223  
GENERAL INFORMATION:  
APPLICANT: Sims, Peter J.  
APPLICANT: Bothwell, Alfred L.M.  
APPLICANT: Elliott, Eileen A.  
APPLICANT: Flavell, Richard A.  
APPLICANT: Madri, Joseph  
APPLICANT: Rollins, Scott  
APPLICANT: Bell, Leonard  
APPLICANT: Squinto, Stephen  
TITLE OF INVENTION: Universal Donor Cells  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kilpatrick & Cody  
STREET: 1100 Peachtree Street, Suite 2800  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30309-4530

```

      FILING DATE: 13-AUG-1996
      CLASSIFICATION: 424
      ATTORNEY/AGENT INFORMATION:
      NAME: Pabst, Patrea L.
      REGISTRATION NUMBER: 31,284
      REFERENCE/DOCKET NUMBER: OMRF 112cipdlv
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (404)-873-8794
      TELEFAX: (404)-873-8795
      INFORMATION FOR SEQ ID NO: 2:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 315 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: cDNA
      US-08-696-777-2

```

Db 76 ACGCATCAAAATCAGATGACACATTGA 50

11 11111 1 1 11 1 11 11

RESULT 14  
US-08-483-433-4/c  
Sequence 4, Application US/08483433  
Patent No. 6100443  
GENERAL INFORMATION:  
APPLICANT: Sims et al.  
TITLE OF INVENTION: Universal Donor Cells  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483.433  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/087,007  
FILING DATE: July 1, 1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/906,394  
FILING DATE: June 29, 1992  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: OMRFL35c1p2 div







2

2



Query Match 30.9%; Score 483; DB 9; Length 483;  
 Best Local Similarity 100.0%; Pred. No. 1,4e-118;  
 Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1016 ctccctggcaggagagcttcagatgttctcttaagatacccggttccaatgctgtatatt 1075  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 483 CTCCTGGCGAGGAGCTTCAGATGTTCTTAAGATACCGGTTCCAAAGTGTGATATT 424  
 OY 1076 ggtttatcaagtttcttcttgaccctcatttaagttctgttcgataagtaacgaag 1135  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 423 GGTATTATCAAGTTTCTTGTGACCTCATTTAAGTTCTGTTCGATAGCATATCAGAA 364  
 OY 1136 gaaaggttatctctggaatttgaaaagttgcctctgttttgattacaagttatcac 1195  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 363 GAAAGTTTATGTCGTGGATTGAAAAGTTGCCCTCTCTGTTTGAATTACAAAGTTATCAC 304  
 OY 1196 acaatacaatcaagttctgtaatacgaagcaacgcatctctgctgataagagacgattc 1255  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 303 ACATCAATCAAAAGTCTGTAATCAGCAAAACACCATGTCGTGATGAGAACGACGATTTC 244  
 OY 1256 ttgcttgctgagagagagagactatattgctgctgagacgtgattaccgaagtaacgtctg 1315  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 243 TTGCTTGCTGCGAGAGGAGACTATATGGCGCTGGAGCGTATACCAAGTACGCTCTG 184  
 OY 1316 agctcttgaggaattgatgaattagagagtgaggaagaatagagatalccattcttatttg 1375  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 183 AGCTTGTAAGATTTGATGATTAAGAGAGTGCAGAAAGATAGATATCCATTCTTTATTG 124  
 OY 1376 taattctgacatgtgtctactcccttgagagaccttgagagcttcttgtagcctgttaac 1435  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 123 TAATTGTGATCATGTTGCTACTCCCTGAGACCTTGAGATGCTCTTTGTACCCCTGTTAAC 64  
 OY 1436 gtccacctgtgacacagtgatatacccttctcgagagatttgttattcttcttattc 1495  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 63 GTCCACCCCTGTACCAAGTATACCTTTCGTGAGATTTGTCTTATTCTCTTACTTC 4  
 OY 1496 aat 1498  
 ||||  
 DB 3 AAT 1

RESULT 2  
 AV560507/c 482 bp mRNA linear EST 07-SEP-2000  
 LOCUS AV560507 Arabidopsis thaliana green siliques Columbia Arabidopsis  
 DEFINITION thaliana cDNA clone SQ135g08F 3', mRNA sequence.

ACCESSION AV560507  
 VERSION AV560507.1 GI:8731933  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
 AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries  
 JOURNAL DNA Res. 7, 175-180 (2000)  
 MEDLINE 20363093  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source 1..482  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="SQ135g08F"

/clone.lib="Arabidopsis thaliana green siliques Columbia"  
 /tissue.type="green siliques"  
 /note="Vector: pBluescriptII SK-; Site1: EcoRI; Site2:  
 XhoI"

BASE COUNT 167 a 109 c 84 g 122 t  
 ORIGIN

Query Match 30.1%; Score 470.8; DB 9; Length 482;  
 Best Local Similarity 98.5%; Pred. No. 2.6e-115;  
 Matches 475; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1026 gggagcttcagatgttctcttaagatacccggttccaatgctgtatattgattatca 1085  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 482 GGAGAGCTTAGATGTTCTTAAGATACCGGTTCCAAAGTGTGATATTGTTATCA 423  
 OY 1086 gttctcttgacctccatttaagttctgttcgataagtaacggaagaagttta 1145  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 422 GTTTTCTTGACCTTCATTTAAGTTTGTTCGATAGTAAATCAGAAAGAGTTTAA 363  
 OY 1146 tgcctggaattgaaagttgcctctgttttgattacaagttatccacaatcaatc 1205  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 362 TGTGTGGAATTGAAAAGTTGCCCTCTCTGTTTGAATTACAAAGTTATCACAAATCAATC 303  
 OY 1206 aaagttgtaatacgaagcaacgcaatgctgtcgtgataagagacgattctgctgtc 1265  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 302 AAGTCTGTAATCAGCAAAACACCATGTCGTGATGAGAACGACGATTCTTCTGCTG 243  
 OY 1266 cgaagacggaactatattgctgctgagacgtgattacaaagtcggtctgagctgtag 1325  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 242 CGAGAGCGGAGACTATATGGCGCTGGAGCGATTAACAAGTAAAGGTTTGTAG 183  
 OY 1326 gaattgataaattagagtgaggaagaatagagatalccattcttattgtaattctg 1385  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 182 GAATGATGATTAAGAGAGTGCAGAAAGATGATATCCATTCTTTAATTGTAATTCGAT 123  
 OY 1386 catgtgtctactccctggaagaccttgagatcttctgtgacgttctgtaacgctcac 1445  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 122 CATGTTGCTACTCCCTGAGACCTTGAGATGCTTTGTGACCTTGTAAAGTACACCTT 63  
 OY 1446 gtacacagtgatacccttctcgagatttgttattcttcttattgtaatacaaga 1505  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 62 GTACACAGCTATACCTCTTGAGATTTGTCTTATTCTTTAGTCAATACACAAG 3

RESULT 3  
 AV519989/c 486 bp mRNA linear EST 01-SEP-2000  
 LOCUS AV519989 Arabidopsis thaliana aboveground organs two to six-week  
 DEFINITION old Arabidopsis thaliana cDNA clone ABD74e07F 3', mRNA sequence.

ACCESSION AV519989  
 VERSION AV519989.1 GI:8679516  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
 AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries  
 JOURNAL DNA Res. 7, 175-180 (2000)  
 MEDLINE 20363093  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/  
Location/Qualifiers

## FEATURES

source

1. 486  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="APD74e07F"  
/clone\_1lb="Arabidopsis thaliana aboveground organs two to six-week old"  
/tissue\_type="aboveground organs"  
/dev\_stage="two to six-week old"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 157 a 101 c 88 g 140 t  
ORIGIN

Query Match 28.8%; Score 450; DB 9; Length 486;  
Best Local Similarity 96.8%; Pred. No. 9.6e-110;  
Matches 459; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 1060 ccaatgctgcatatttggtttatcaagtttcttctgacctcatttaagttctgtcg 1119  
DB 486 CCAATGCTGATATTGGTTATCAAGTTTCTTGACCTCATTTAAGTTCTGTTG 427  
OY 1120 ataggaatcaagaagaagtttattgtctggatttgaagaattgcctctgtttg 1179  
DB 426 ATAGGAATCAGAGAGAAAGTTTATGTTGGATTGAAAAGTTGCCCTCTGTTTG 367  
OY 1180 attacaagaattacacacatcaatcaaatcgttgaatcaggaacagcatgtctgc 1239  
DB 366 ATTACAAAGTTATCACACATCAATCAAGTCTGTATCAGGAAACAGCATGCTGTC 307  
OY 1240 gatggaagcagatctctgtctgctgagagagagacatatgctgtgagcgtgatt 1299  
DB 306 GATGGAAGCAGCATCTCTCTGCTGCGAGAGCGCATATATGCGCGTGAACGTGATT 247  
OY 1300 accaagtagcgtctgagagcttgaagaattgaatgaagaatcgaagaatgaat 1359  
DB 246 ACCAAGTAGCGGTCTAAGCTGTGAGATTGATGAGAGTGCAGAAATATGAT 187  
OY 1360 atccattcttatttgaattctgacatgttctacccctgagaccttgagatgctct 1419  
DB 186 ATCCATTCTTTTATGTATTTCTGATCATGTGCTACTCCCTGAGACCTTGATGCTCT 127  
OY 1420 ttgtgacctgttaacgctccaccttgtaaccaagtgataacccttctggagatttgc 1479  
DB 126 TTGTAGCCCTGTGAACGTCACCCCTGTACACAGTGTATACCCCTTAAAGAAGCCCTGTG 67  
OY 1480 ctattctcttgaatcaacacagaagcgtgtatccctgagagcttattgcaagaa 1533  
DB 66 CTTATTCTCTTATGTTCAATACAGAAAGTATCTGTGAGCTTTATTGCAAAAA 13

## RESULT

4

BM359804 885 bp mRNA linear EST 09-JAN-2002  
LOCUS

DEFINITION GA\_Ea0023003r Gossypium arboreum 7-10 dpa fiber library Gossypium  
arboresum cDNA clone GA\_Ea0023003r, mRNA sequence.

ACCESSION BM359804  
VERSION BM359804.1 GI:18100550

KEYWORDS EST.

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum

REFERENCE

AUTHORS

TITLE

JOURNAL

## COMMENT

Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Total High Quality bases = 321  
Seq primer: TAATAGACTCTACTATAGG  
High quality sequence start: 3  
High quality sequence stop: 524.  
Location/Qualifiers

## FEATURES

source

1. 885

/organism="Gossypium arboreum"  
/strain="AKA"  
/cultivar="8400"  
/db\_xref="taxon:29729"  
/clone="GA\_Ea0023003r"  
/clone\_1lb="Gossypium arboreum 7-10 dpa fiber library"  
/tissue\_type="fibers isolated from bolls harvested 7-10 dpa"  
/lab\_host="E. coli"  
/note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 258 a 166 c 195 g 266 t  
ORIGIN

Query Match 27.5%; Score 430; DB 10; Length 885;  
Best Local Similarity 72.3%; Pred. No. 2.7e-104;  
Matches 599; Conservative 0; Mismatches 225; Indels 4; Gaps 3;

OY 377 ttactctgtacaatttctctcgcagatggtgacctatcagatctgaatcctatgctgag 436  
DB 1 TGACGGTATACCAATGCTCTGAGAGAGGTGTATACCTGTTTGCAAGTATGTTGATG 60  
OY 437 aagataagaagaagctgtttacacgtaagttggcgctgttgaatggaaaccat 496  
DB 61 AAGATAAGATGAGTCTTCTTCTACCTGTAAGTTGGGCTGCACATTTGATGATCACAT 120  
OY 497 atgttcgcgtgtaggaagaataagataatccgaatcattgaacagtgaaagca 556  
DB 121 TTGTGTTGCTGAGAGTATTAATGATATCCGTGTAATGATGAGCAATGAGAAAA 180  
OY 557 ttcataagagctctgtggttcaatgagatcagttgaagaagaatcagacaacattta 616  
DB 181 TACCAAGAGATTTGTTGGACATGATCTTAACGAAACACGATCAACCACTGA 240  
OY 617 aacctcaacttgatcattgataattgctgagcctgagagtcacgtataagaattcagtg 676  
DB 241 AACCATCACTTATAGTGTCTGCAAGCAAGATATATCGGTTAGATTGTGATGTTCA 300  
OY 677 ctggagatagtatttgaatattgctgagcctgagagtcacgtataagaattcagtg 736  
DB 301 CTGGAATCTGCAATTTGATTTGCTGAGACCGGGGACACCCGATGAAGTTTGAGTG 360  
OY 737 tggatttcatcctgctgtatatttaccgcttctgctggtgtgtatggaaacacattta 796  
DB 361 TGGATTTCATCTTGAAGCATTTATGCAATTCAGATTGCCGATGAGAACACGTGA 420  
OY 797 aatatgtcaatgaagaagatttggagctagctgagagtcacgtatcacaatgagcag 856  
DB 421 AGATTGGTCAATGAAGAAGATTGTGACATTATGTGAGAAATATATTCACATGACAGATC 480  
OY 857 atccatcaaatcccccacacaaattgttccaattccctglatattacagcttccatcata 916  
DB 481 TTCTTACAAATTTCCAAACATTAAGCTACAGTACCTGCTGCTTACGATCAATACAT 540  
OY 917 caaatatgtagatttgaacgctgtgtgtgtgtatatttaccctcacaagaagtgtag 976  
DB 541 CAATATATGAGCTGATATCTGTGACCTGATTTATTCATTTGTCMAAGAGTGTGAC 600  
OY 977 acgagatccgtgtgtggagaccacactgaagaagaattccctgagagagctgag 1036

[illegible]

```

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/notes="vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT      261 a      203 c      263 g      259 t      1 others
ORIGIN

Query Match      23.6%; Score 369.4; DB 10; Length 987;
Best Local Similarity 72.5%; Pred. No. 4.9e-86;
Matches 492; Conservative 0; Mismatches 186; Indels 1; Gaps 1;

OY 400 gatgtgcatatcagcatcgaatcctatctgataagaagaagatcgctttac 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 96 GAGGTGTATAGCTGTTTGCAGCTTATGTGATGAAGATAGATGCTTCTAC 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 460 acggtgaagtgtggcggtgtggttaattgggaaccatagtgcggctggaggataaa 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 156 ACTGTAGTGTGGCCGCGACATGTGATGTCACCATTTGTTGTTGAGTATAT 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 520 ggtataacacagatcagtcagtcagtcagtcagtcagtcagtcagtcagtc 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 216 GGCATTAATCCGTTATGATGTCGAGCATGGAATAATACACAAGATTTGTTT 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 580 ggaagatcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 276 GGTGACTGTATTAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 640 agcaagatgaatctgtctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 336 AGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 700 gctggaagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 396 GCTGAGCGCGGGGACACCGCATAGATGATGATGATGATGATGATGATGATG 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 760 taccgtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 456 TATGCAATTTGCAAGTTGGGGATGAGCAACACTGTAAGATTTGCTCAATGA 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 820 tggagctacgtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 516 TGGACTTATGTGAGAAATATTCATGAGCGGATCTTACTTCAAAATTTCCACA 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 880 ttgtgcaatccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 939
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 576 TACGTACATTTCCCTGTGTCTTACCTTCAGTTCAATTAATGATGAGACTGAT 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 940 tggtttgatattatctctcaagaagtgtggaacagagatcctgttggaacca 999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 636 TGGCTTGTGATTTTCAATCTGTCAAGAGTGTGTGCTACCAATATCTGCTAT 695
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1000 caactgaagaagaatctctctgtgagaggagtcacagatgtcttaataag 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 636 AAGATTAAGGAGCA-TGTGCAAGAGAGGAGACAGCTGACATCTTGAAGAAG 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1060 ccaatgtgtgatatgtgt 1078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 755 CCGAATTTCGACATTGGTT 773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
BI424774 585 bp mRNA linear EST 29-NOV-2001
LOCUS
DEFINITION
ID: Gm-c1036-4803 5' similar to TR:Q9XFA4 Q9XFA4
FERTILIZATION-INDEPENDENT ENDOSPERM PROTEIN.; mRNA sequence.

```

```

ACCESSION BI424774
VERSION BI424774.1 GI:15201253
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 585)
REFERENCE Shoemaker,R., Kelm,P., Vodkin,L., Expelting,J., Corryell,V., Khanna
AUTHORS A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Waterston,R. and Willson,R.
Public Soybean EST Project
Unpublished (1999)
JOURNAL Contact: Shoemaker R/Public Soybean EST Project
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccl@resgen.com
High quality sequence stop: 433.
FEATURES
source
1..585
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone_lib="Gm-c1036"
/clone_lib="Gm-c1036"
/tissue_type="Somatic embryos cultured on MSD 20"
/lab_host="DH10B"
/notes="vector: pSPOR1; Site_1: NotI; Site_2: SalI. This
cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies SuperScript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dt) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPOR1 vector. The ligated cDNA fragments
were transformed into E. coli Electromax DH10B host cells.
This library was constructed in the laboratory of Dr. Lila
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

BASE COUNT      167 a      97 c      127 g      194 t
ORIGIN

Query Match      23.2%; Score 363.2; DB 10; Length 585;
Best Local Similarity 76.4%; Pred. No. 1.9e-86;
Matches 446; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

OY 643 aaggatgaatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 AAGATGAATCTGTCCGGTATGATGATGATGATGATGATGATGATGATGATG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 703 ggaagtcgaagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 GGAGCTGGAGGACATCGTATGATGATGATGATGATGATGATGATGATGATGAT 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 763 cgcttgctagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 CGTATTGCTAGTGTGGCAGTATGATGATGATGATGATGATGATGATGATG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 823 acgtacgtcgaagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 181 ACATATGAGAAAATCATTTACATGACACTGATCTTCAAGTTTCCCAAAATAT 240  
 QY 883 gtccatccctgttattcaacgtccatccatcaataatgtagatgtaacgttgg 942  
 Db 241 GTCCAGTTTCTGTTACAAAGTCTTCAAGTCTTCAATTAATGTTAGTACGTAATAGGAG 300  
 QY 943 ttgtgtatttattcctcctcaagaagtgtagaacaagagatcctgttggagaccaca 1002  
 Db 301 CTAGTGATTTTATCTCTCAAGAGTGTGACAAATGAATATCTTGGGGAACCTAA 360  
 QY 1003 ctgaagaagaattcctcctggcgaagagagcttcaagatgtctcttaagatacccggttcca 1062  
 Db 361 GTGAAGGAACAACTCCAGGGGAGGCTGATGTTGACATCTTCAGAAATACCTGTCTT 420  
 QY 1063 atgtgtatttattgttattcaagatttctgtgacctccattaaagtctgttgcgata 1122  
 Db 421 GAGTGTGATATCTGTTCTCAAGTCTTCTTGTGATTTCCATTTCACACTTACCGCAGTG 480  
 QY 1123 ggtatcaagaagaaggttattatgtctgtggaattgaaaaagtgcctcctgttatt 1182  
 Db 481 GGTAAAGGGAAGGAGGATTTTGTGGAATTGACATGCAAGTCTCTCTGACTTGT 540  
 QY 1183 acaaatgtatcaacaatcaatcaaatctgttaattcaggcaaac 1226  
 Db 541 GCAAGGTGTGACATCTCAATCAAAATCTCCATCAGCAGCAGAC 584  
 RESULT 8  
 BE888279 658 bp mRNA linear EST 30-MAY-2001  
 LOCUS EST514130 cSTD Solanum tuberosum cDNA clone cSTD9B13 5' sequence.  
 DEFINITION mRNA sequence.  
 ACCESSION BE888279  
 VERSION BE888279.1 GI:14265365  
 KEYWORDS EST.  
 SOURCE potato.  
 ORGANISM Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 REFERENCE 1 (bases 1 to 658) Bezzerides, J., Ewing, E., Cho, J., Chieningo, A.,  
 Bougri, O., Buel, C.R., Ronning, C., Tanksley, S. and Baker, B.  
 Generations of ESTs from dormant potato tubers  
 Unpublished (2001)  
 TITLE Contact: Cathy Ronning  
 JOURNAL The Institute for Genomic Research  
 COMMENT For clone info: please contact Research Genetics, Libraries  
 Division tel 1-800-711-6195, email cdna@resgen.com  
 Seq primer: M13p-R.  
 FEATURES  
 source  
 location/Qualifiers  
 1..658  
 /organism="Solanum tuberosum"  
 /cultivar="Kennebec"  
 /db\_xref="taxon:4113"  
 /clone="cSTD9B13"  
 /clone\_1bp="cSTD"  
 /tissue\_type="dormant tuber"  
 /dev\_stage="one month post-harvest"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; This library targets genes expressed in dormant  
 tubers. This library was made from sections of dormant  
 tuber, avoiding the buds and epidermis. Tubers were stored  
 for one month post-harvest at 4°C. The tuber was peeled,  
 well away from the surface. Then it was chopped into 1-2  
 mm cubes and immediately frozen in liquid nitrogen. This  
 library is noted as p4 in Tanksley lab notebooks."

Query Match 22.8%; Score 356.8; DB 10; Length 658;  
 Best Local Similarity 74.7%; Pred. No. 1e-84;  
 Matches 487; Conservative 0; Mismatches 162; Indels 3; Gaps 3;

QY 301 cctttagtgcgtgttcttcaactccttgatgctgcttcttgatgctgttcc 360  
 Db 8 CCATTATATGGCGTCTTCAACTTCAATGATTTGGCGTACTTCAATGCTTCTTACT 67  
 QY 361 gctgttgaaatcgcgattcctctgtacaaattgtctcga-gatgtgcatacagcatt 419  
 Db 68 GTGGCGGGAATCGAGTCACTGTATATCATGTCGGAAGGTGGTGTATGCGTGT 127  
 QY 420 gcaatccatgtctgtatgaagaagaag-gagtcgttttacacggtlaagtgtggcgttg 478  
 Db 128 GCAGCTTATATTGATGAAGATTAAGAAATGCAATCCCTTTTACACTGTAAGTTGGCGTCA 187  
 QY 479 gcgttaatggaaaccatagttgcgcgtgagagtaaaagtataatccgaagcatg 538  
 Db 188 ATATTGATGGAGTCCATCTTAGTGCTGTGGAATTAATGAGTTATCTGTATTG 247  
 QY 539 acgtcaacagtgaaacagatcataagagctgtgtgtgcatgagatcagtgaaacaa 598  
 Db 248 ATGCTGGGAAGAGAGATPACACAGAGCTTGTGGGACACGAGACTCAGTAATAAGAA 307  
 QY 599 tcaagacacaccttaaaacctcaactgtgattactgtctagcaagatgaatctgttc 658  
 Db 308 TTAGGACTCAACCTTGAAACCATCTTGTATTATCTGCACCAAGATGATCTGTTC 367  
 QY 659 gttctgtgaatgttgaactggagatgattgattgatttctgtgagctgtgagctc 718  
 Db 368 GCTTGTGGAATGTTCAATGAGATGCAATTTGCTATTGCTGTGGTGGGCTCAGC 427  
 QY 719 gctataagttcctaagtgtgagatttcaatccgcgtgatattacacgcttgcagtg 778  
 Db 428 GGAATGAGTACTTGAAGTGGAGCTTCCATCTCTCATATATTCGGATTGCAAGCTG 487  
 QY 779 gtagaacaccactataaaatattgtcaatgaagaagtttggagctgtgagaagt 838  
 Db 488 GAATGATTAACACTGTTAAGATCTGTCATGAAGAAATCTGACATATGAGAGAAAT 547  
 QY 839 cattcacatgactgatatgcatcaaatccccaataattgtccattcccgat 898  
 Db 548 CGTTTACTTGAGGAGATCTTCTTCAAGTTTCCCATTAATATGATACATTTCCATTGT 607  
 QY 899 ttacagttcattcattacaataattgtagattgtaacgctgtgtgtgta 950  
 Db 608 TAATAGCTTCTGTCCATAC-AACTACGTTGACTATACATGATGCTGTGTA 658  
 RESULT 9  
 BE353166 624 bp mRNA linear EST 18-MAY-2001  
 LOCUS EST340387 tomato mixed elicitor, BRT Lycopersicon esculentum cDNA  
 DEFINITION clone cLET41E1, mRNA sequence.  
 ACCESSION BE353166  
 VERSION BE353166.1 GI:9291142  
 KEYWORDS EST.  
 SOURCE tomato.  
 ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 REFERENCE 1 (bases 1 to 624)  
 AUTHORS D'Ascenzo, M., He, X., Lyman, J., Matern, A.L., Viston, T., Holt, I.E.,  
 Liang, F., Hansen, T.S., Ronning, C.M., Craven, M.B., Bowman, C.L.,  
 Nierman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D., Giovannoni,  
 J.J. and Martin, G.B.  
 Generation of ESTs from tomato callus (mixed elicitor)  
 Unpublished (1999)  
 TITLE Contact: CUGI  
 JOURNAL Clemson University Genomics Institute  
 COMMENT Clemson University

BASE COUNT  
 ORIGIN

174 a 112 c 159 g 213 t



100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

## FEATURES

source

location/Qualifiers

1. 624

/organism="Lycopersicon esculentum"

/cultivar="Rio Grande Ptor"

/db\_xref="taxon:4081"

/clone="CLET41E1"

/clone\_lib="tomato mixed elicitor, BTI"

/cissue\_type="leaf"

/dev\_stage="4-6 week old plants"

/lab\_host="X11-Blue MP"

/note="Vector: pBluescript SK(+); Site.1: EcoRI; site.2: XhoI; cLET - Inoculated with a variety of disease response

elicitors. Plants exposed to 2,6 dichloroisonicotinic

acid, BTX, jasmonic acid, ethylene, fenthion, ETX,

okadaic acid, or systemin prior to tissue harvest. EcoRI

site was destroyed during cloning."

BASE COUNT 172 a 107 c 160 g 185 t

ORIGIN

Query Match 22.2%; Score 346.6; DB 10; Length 624;  
Best Local Similarity 73.1%; Pred. No. 5.5e-82;  
Matches 445; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 209 taacttaggaagcaagtaataagttggcttctgactccatgaataagaatcgtaca 268  
DB 16 TACCTTAGAGAGTGTAGCGAGTGTGATTAACACCGCTCAAGAAAGAGCTATA 75  
QY 269 aagtagcaagatagattcgaagaagaacacctgtagctgtgttttctacttc 328  
DB 76 GAGTAGCAGACAGACTCCAGAGAGCAAGCCCATTAATAGCGCTTCAATTCA 135  
QY 329 ttgagtcgtcttctcgtatgcttctgacgcgtggtggaatcggatctgtaca 388  
DB 136 TTGATTGCGGCTACTTCAATGCTTGTACTAGTGGCGGAATCGTGACGTATAC 195  
QY 389 atgtctcggagatggtgcacatacgaacttgcacatcctatgctgataagaagaag 448  
DB 196 AGTGTCTCAAGGTGTGTATGCTGTCTGACAGTCTTATATTGTAAGATAAGATG 255  
QY 449 agtcgtttacacgtaagtgtggcggtggtgtaattggaaccatattgtggcgctg 508  
DB 256 AATCTTTTACCTGTAAGTTGGGCTGCAATATGATGAGAGTCCATTCTTAGTGCTG 315  
QY 509 gaggagtaaaaggtatatacgcagtcattgacgtcaacagtgaaacgattcataagatc 568  
DB 316 GTGGAATAATGAGATTATTCGTATGATGCTGGAGAAAGAGATACCAAGAGCT 375  
QY 569 ttgtgtgataaggaattcagtaagaacaaacgaacacacacacacacacacacacac 628  
DB 376 TTGTGGGGGACGACGACTGATAAAGAAATAGGACCAACCTTGAAACCATCTCTCTG 435  
QY 629 tgattacgtcaggaagaagtaactgttctgttctgttgaatttgaacttgaattgtga 688  
DB 436 TTTTATCTCCAGCAAGATGATCTGTTGCTTGTGGAATGTTCTACTGGGAATATGA 495  
QY 689 ttltgataattgctgagatctgagtcacgcatacgaattcctaaattcgtgattcacc 748  
DB 496 TTTTGTATTCTGCTGGCGCTGGAGGTCATCGGAATGAAATGACTACTGTGACTTCATC 555  
QY 749 cgtctgataattacgcgttctgtagtgtgtgtatgagacacacacacacacacacacac 808  
DB 556 CTACGCAATATATGAGATTGCTAGCTGTGAAATGATTAACCTGTTAAGATCTGGTCA 615  
QY 809 tgaagaagt 817  
DB 616 TGAAGAAT 624

RESULT 10

BE918589

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 622)

REFERENCE

AUTHORS

L.H.

JOURNAL

COMMENT

An EST database from Sorghum: ovaries of varying immature stages

Unpublished (2000)

Contact: Cordoniier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1805

Fax: 706 542 1805

Email: mmprratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions

below phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: JEN REV

High quality sequence stop: 530

POLYA=NO.

FEATURES

source

location/Qualifiers

1. 622

/organism="Sorghum bicolor"

/db\_xref="taxon:4558"

/clone\_lib="Ovary 1 (OV1)"

/note="Organ: Mix of ovaries of varying immature stages

from 8-week-old plants; Vector: pBluescript II from lambda

zap II; Site.1: XhoI; Site.2: EcoRI; The library was made

from poly-A RNA in the cloning vector lambda Zap II.

Clones to be sequenced were prepared by mass excision."

BASE COUNT 182 a 114 c 147 g 179 t

ORIGIN

Query Match 21.8%; Score 341; DB 10; Length 622;  
Best Local Similarity 72.7%; Pred. No. 1.7e-80;  
Matches 440; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 550 gaaacgattacataagagctgtgtgggtcagtgagattcagtgaaacgaacacagacacaa 609  
DB 14 GACAAATTAGCTAAGAGCTTTGTGGCCATGCTGACTCAATAAATGATAGAACTCA 73  
QY 610 cctttaaaccctcaacttctgattactcagtagaagaatgaatctgtctgttggaaat 669  
DB 74 CCGTTGAAGCCTTGGCTCATCATTTCTGCAAGCAAGATGAATCTTTGGCATGGAAT 133  
QY 670 gttgaactcggagatgatatttgataattctcgtgagcgtgagagtcacgcataagatc 729  
DB 134 GTCCATACAGGGATCTGTGATTTGATTTGCTGAGAGCTGAGAGTCATGCAATGAAGTG 193  
QY 730 ctaagtgatttattatccgcctgatttaccgccttggtagtgatgtagacac 789  
DB 194 TTGAGTGTGACTTTCACCTTACCTGATGATGAGAGCTGTTCAGAGCTGTGCAATGGA 253  
QY 790 actatnaaatatgylcaatgaaagatttggagacagtlagcagagaagtcacatcagatg 849  
DB 254 ACTGTGAATAATCTGTCATGAGAAAGATTGCTATATGTTGACAAATCATATTCATG 313  
QY 850 actgatgatccataaattcccaacaaatttgcacattccctgtatttaccgctcc 909  
DB 314 ACTGACCTTCCATGAGAGTTTCCACAAAATATATGTCAGTTTCGGGCTTGATTGCTGCA 373  
QY 910 attcatacaaatatgtagatgtaaccgltgtgttgtagatttattccctcacaagaagt 969

Db	374	GTACCTCTACTATTATGTGACTGTATCAAGATGGCTTGCTGACTTCATCTCTATCAAAAGAGT	433
Oy	970	gtgaacacacagatctcctbctgttggaaccaccaacctgaaagaatctccctggcgaagga	1029
Db	434	GTTGACAAATGAAATTGTCTCTGTGGACCAACAAAGACAAAGACAGAGTCCTGTGGGAGGGA	493
Oy	1030	gcttcgaatgtctatattaagatacccggttcccaatgctgatalattggtttatcaagttc	1089
Db	494	AGCATCGATTATCTTCAAAAGATACCCTCTCCACGATGTGACATCTGGTATCAAAATTT	553
Oy	1090	tctttgacccccaatttaagcttcgttgcgataagtglaacggaagaagttatgtc	1149
Db	554	TCATGTGATTTTCACTTCATCATCATGCTGTGTCATAGCAGATCGGAAGGAATAATATACGTG	613
Oy	1150	ttggga	1154
Db	614	TGGGA	618

[illegible]

ORGANISM	REFERENCE
Solanum tuberosum Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Aspermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Spermatidae; euasterids I; Solanales; Solanaceae; Solanum. 1 (bases 1 to 578)	van der Hoeven, R.S., Bezzerides, J., Holt, I.E., Liang, F., Cho, J., Uttérback, T., Hansen, C.L., Doan, B., Bougri, O., Buell, C.R., Rønning, C.M., Fry, W.E., Tanksley, S.D. and Baker, B. Generation of ESTs from potato leaves and petioles unpublished (2000)
TITLE JOURNAL COMMENT	Contact: Cathy Rønning The Institute for Genomic Research For clone request: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdnatresgen.com.

FEATURES	Location/Qualifiers
source	1. .578

/organism="Solanum tuberosum"  
 /cultivar="Kennebec"  
 /db\_xref="taxon:4113"  
 /clone="CSTB2P18"  
 /clone\_1lb="potato leaves and petioles"  
 /tissue.type="leaflets and petioles"  
 /dev\_stage="8 weeks old plants"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; Tissue was supplied by Dr. Fry (Cornell University),  
 leaflets and petioles were isolated from 8 week old  
 greenhouse grown plants. The plants were watered and  
 fertilized freely. The tissue was immediately frozen in  
 liquid nitrogen."

Query Match	20.48;	Score 319;	DB 10;	Length 578;
Best Local Similarity	72.28;	Pred. No. 1.3e-74;		
Matches 415; Conservative	0;	Mismatches 160;	Indels 0;	Gaps 0;

[illegible]

Db	64	AGAGATATGAGCGACGAACACACTCCAAAGGCAAGGCCAAAGGCCATTAATATGCGTCGTCCTT	123
Oy	321	caacttccctatgctgcgtttcttcogatgtaattgtaaccgcctgtgtgaatctgatctac	380
Db	124	CAACTTCATTGATTCGGCCCTACTTCAAGTCTCTTGCTACTGTGGCGGAATTCGAGTAC	183
Oy	381	ctgtacatctgtctcgcgagatgtygtccatactgacatctgaatccctatgctatbgaaga	440
Db	184	TGTATATCAGTGTCGCAAGGCGTGATTATGTCGTGCTGCAGTCTTAATATGATGAGAGA	243
Oy	441	taaggaagatgcgttttcaacggtgaagttggcgtgtgtgcgttaattggaaaccataagt	500
Db	244	TAAAGATGAATCCTTTTCACCTGTAAGTTGGGCGCCGCATATGTGATGGGAGTCCTCTT	303
Oy	501	tgcgcctggagagatgaagaagtataatcccgatctatgactlcaacagtgaaacgattca	560
Db	304	ACTGCGTGTGGAAATAAATGAGATTAATTCGTATTATGATGCTGGGAAGAAGATACAA	363
Oy	561	taagagctctgtgggtcatgtgaatcaagtgaagaagaatcggacacaacctttaaac	620
Db	364	CAAGAGCTTTTGGGGCGACGGAGCTCGATGAATATAATTAATTAAGGACTCAACCTTGAAACC	423
Oy	621	tcaactttgattactgctacgaagaagatgaaatctgttcgtttgtgtgaatgttgaacctg	680
Db	424	ATCTCTTGATTATTCGCCAGCAAGAGATATCTGTTCCTTGTGGAAATGTTCAATACGG	483
Oy	681	gatactgatatgtgaatatctgtcgtgaacgttgaagtcacgcgtatgaagttctaagtgtgga	740
Db	484	AATATGACATTTTGGATTTGCTGCGTGGCGAGAGTCATGCGAATGAAGTACTAGTGTGA	543
Oy	741	tttcatccgtctgatatctaacgctttgcaagt	775
Db	544	CTTCCATCTCTACTGATATATATTCGATTCGATTGCTTACCT	578

RESULT	12
BF256342	
LOCUS	
DEFINITION	BF256342 773 bp mRNA linear EST 22-OCT-2001 HVSNEIF009E24f Hordeum vulgare seedling root EST library HVCDNA0007 (Etisolated and unstrressed) Hordeum vulgare cdna clone HVSNEI009E24f, mRNA sequence.

ACCESSION	BF256342
VERSION	BF256342.2
KEYWORDS	GI:13118123
SOURCE	EST
ORGANISM	barley.
	<i>Hordeum vulgare</i>

REFERENCE  
AUTHORS  
1  
Wing, R., Close, T.J., Kleinof, A., Wise, R., Begum, D., Fritsch, D., Yu  
1 (bases 1 to 773)  
; Trilicaceae; Hordeum.  
Euarystia; Viridiplantae; Streptophyta; Embryophyta; tracheophyta  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
; Trilicaceae; Hordeum.  
Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Chol, D.W., Penton  
; R.D., Oates, R. and Main, D.  
Development of a genetically and physically anchored EST resource  
for barley genomics: Morex unstressed seedling root cDNA library  
Unpublished (2001).  
On Nov 16, 2000 this sequence version replaced gi:11185371.

Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: [twing@clemson.edu](mailto:twing@clemson.edu)  
Total hg bases = 545  
Seq primer: AATTAACCTCACAATAAGG  
High quality sequence stop: 695.  
Location/Qualifiers  
1..773

```

/organism="Hordium vulgare-
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HvSMEf0009E24f"

```

/clone.lib="Hordeum vulgare seedling root EST library  
HVCNDA0007 (Etiolated and unstressed)"  
/tissue\_type="Seedling root"  
/lab\_host="TUC121"  
/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI;  
Seeds were surface sterilized then germinated under axenic  
conditions in the dark at room temperature on filter paper  
with water. Nystatin and cefotaxime in covered  
crystallization dishes. Five-day old seedling roots were  
then harvested, total RNA was prepared, poly(A) RNA was  
purified, one primary unamplified cDNA library was made,  
and 1 million pfu were in vivo excised to give Bluescript  
SK(-) cDNA phagemids. These steps were performed in the TU  
Close laboratory at the University of California,  
Riverside (Choi, Close, Fenton). Phagemids were plated and  
picked at the Clemson University Genomics Institute (CUGI)  
(Begum, Palmer, Frisch, Atkins and Wang). Plasmid DNA  
preparations, DNA sequencing and sequence analysis were  
performed at CUGI (Wang, Yu, Frisch, Henry, Simmons, Oates  
Rambo, Main). The sequence has been trimmed to remove  
vector sequence and contains a minimum of 100 bases of  
phred value 20 or above. For more details on library  
preparation and sequence analysis see  
<http://www.genome.clemson.edu/projects/barley>. To order  
see Close TJ, Wang R, Kleinof A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

BASE COUNT 210 a 158 c 195 g 210 t  
ORIGIN

Query Match 20.3%; Score 317.2; DB 10; Length 773;  
Best Local Similarity 70.0%; Pred. No. 4.5e-74;  
Matches 441; Conservative 0; Mismatches 188; Indels 1; Gaps 1;

OY 666 gaatttgaacgtgagatgatttgaatttctgagagtcgagtcacgtcatga 725  
DB 1 GAAATGCCATTAAGGAGATCGATTTGATTTTCTGGAGAGGCGACCGTAAATGA 60

OY 726 agttcaagtgtgattttcacgcgtctatattaccgcttgctgagtgtgataga 785  
DB 61 AGTATTGAGTGTGACTTCCACCTTGTGATATCAACCAATTTGCCAGTTGGCATGGA 120

OY 786 caaccatataaataatgtaaaagatttggagctgacgtcgagaagatcattcac 845  
DB 121 TAAATCTGTTAAATCTGCTCAATGAAGAAATTTGGCCATCGTGAGCAAAATCTTTAC 180

OY 846 atggactgattgattcataaaattcccaaaaatttgcacattccctgtatttaagc 905  
DB 181 ATGGACTGACCTTCATCAAAATTTCCAAATATATGTCCAATTTCCACTATGACTTC 240

OY 906 ttccattcatacaaatatgtagattgtaacgcttggttggatttattctctccaa 965  
DB 241 CGTGGTGCAATTTACTACTATGACTGATAGTGCGTGTGATGACTTCACTCTGCGAA 300

OY 966 gagtgtagaacaagatcctgtgtggaaccacacaactgaagaagaattctctggca 1025  
DB 301 GAGTGTGACAAATGAATTTGTTCTGGGAGCCAAAACAAAGAACAGGCGCCGAGA 360

OY 1026 gggagcttcagatgttctatttaagataccggttccaagtgtgattattgtttatcaa 1085  
DB 361 GGGTAGCAATTAATGCTTCAGAAAGTACCTGTGCCGATTTGACATTTGGTTATACAA 420

OY 1086 gttctcttgtagccctcatttaagtcgtgtgagataagtaacgaagaagaagtta 1145  
DB 421 ATTCATGATGATTTTCACTTCATCAATTAAGCAATGAGCAACCGGAAGGCAAAATCTGA 480

OY 1146 tgtctggagatttgaaagttgcccctccgttttgaattacaagtatacacaaatcaac 1205  
DB 481 TGTGTGGGAAGTGACAGAGCCCTTCGTGCTAATTTACCGCGTGTAGTAGTCCGCAATG 540

OY 1206 aaagctgtatcagcaaacagccatctgtctgtagtgaacacagattctgtctgc 1265  
DB 541 CAAATGCCAAATTAAGCAACATGCTGCTTGTGAGACACATCTGCTCGCG 600

OY 1266 cgaagacgg-gaactatgagcgtggagc 1294  
DB 601 GGAGATGCGCAACATATACCGCTGGGAG 630

## RESULT 13

AM223682 613 bp mRNA linear EST 18-MAY-2001  
LOCUS  
DEFINITION EST300493 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA  
clone cLEM12D2, mRNA sequence.

ACCESSION AM223682  
VERSION AM223682.1 GI:6535366  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 613)

Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Jiang,F.,  
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Rønning,C.M.,  
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.  
Generation of ESTs from tomato fruit tissue  
Unpublished (1999)  
Contact: CUGI

## JOURNAL

## COMMENT

Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
Location/Qualifiers  
1..613

## source

/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEM12D2"  
/clone.lib="tomato fruit red ripe, TAMU"  
/tissue\_type="pericarp"  
/dev\_stage="red ripe (7-20 days post-breaker)"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Giovannoni; Fruit were tagged at the  
breaker stage (first sign of lycopen accumulation on the  
blossom end of the fruit) and harvested 7 days  
post-breaker (fully red-ripe), 10 days post breaker, and  
20 days post-breaker (over-ripe). 20 day fruit which  
showed external or internal signs of pathogenesis were  
discarded. Fruit were cut in half and the seeds and  
locules were discarded prior to freezing the pericarp."

BASE COUNT 186 a 108 c 137 g 182 t  
ORIGIN

Query Match 20.0%; Score 312.2; DB 9; Length 613;  
Best Local Similarity 71.6%; Pred. No. 9.1e-73;  
Matches 410; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

OY 731 taagtgtgattttacatccgtctgataattacgcgttctgctggtgtgtatgagacaca 790  
DB 1 TTAAGGGGAGACTTCCTACTGATATATTCGATTTCTGCTGCTGTGGAATGATTAACA 60

OY 791 ctattaaatggtuacatgaagaagtttggagcgtgagcgtgaggaagtcattcacatga 850  
DB 61 CGTTTAAGATCGTGTGTAATGAAGAATTTGTGACATATGTGAGAAATCGTTACTTGA 120

OY 851 ctgatgatccatcaaaattcccaaaaattgtccaattccctgtatttaccgttcca 910  
DB 121 CGGATCTTCCTTCCCAAGTTTCCACAAAATATGTATACAGTTTCCATTTGTAATAGCTTCG 180

Qy	911	ttctacaactatgatgatgtatgtaacccgttgatttggatatttatactctccaaagggtg	970
Db	181	ttccattacaaactgtacgttgactgtataacagatggccttgggtgattttatcttgcacaaagtcg	240
Qy	971	tgcagcaacgagatccctcgttgttgggaaccacaacactgaaagagatctccctcgcgagggag	1030
Db	241	tttgacaaatgaaatttcatttatatgggacaccaaagatgaaagaaacaaatctcctcgagagcgca	300
Qy	1031	ctcagaatgctctatataagaatacccggttccaatgctgataatttgatttcaagttc	1090
Db	301	ccagagagactcctccaaagaatgatcctctggccggagtgatgatatgtgttttttaagcgttt	360
Qy	1091	cttgtgacctccatttaagtcttgcttggtgatagataacagaaagaaaggtttatgctt	1150
Db	361	catgttgattttacactacacaaacagagcgttatatgggaataatgacgaagcacaagatctttgtgctt	420
Qy	1151	ggagattgaaaaagtgccctccctcgttcttgatatacaaaagtataccacacataacatcaagaat	1210
Db	421	gggaaggtacaaacaaagcccgcccaactttgattgacaaacgctgctcatatgttcaatctaamc	480
Qy	1211	ctgtaatcaagcnaaacagcaatgctctcgatgtagaagcagatctctgtgcgcgaggg	1270
Db	481	aaccattatgacacagactcccatgctcttttgatgtagaagcaccataactatgttcgttctgtaag	540
Qy	1271	acggagcatatagcgctgggagcgtgattacaa	1303
Db	541	atgggactatataatggcgctgggaatggtagacaa	573

RESULT	14
AV914549	
LOCUS	608 bp mRNA linear EST 18-JAN-2002
DEFINITION	AV914549 K. Sato unpublished cDNA library, cv. Haruna Nijo germination shoots Hordeum vulgare subsp. vulgare cDNA clone bags608 5', mRNA sequence.
ACCESSION	AV914549
VERSION	AV914549.1 GI:18210326
KEYWORDS	EST.
SOURCE	Hordeum vulgare subsp. vulgare
ORGANISM	Hordeum vulgare subsp. vulgare

REFERENCE 1 (bases 1 to 608)  
AUTHORS Sato, K., Saitoh, D. and Takeda, K.  
TITLE Barley EST sequencing project in NIG and Okayama Univ.  
JOURNAL unpublished (2002)  
COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel.: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: [tshini@genes.nig.ac.jp](mailto:tshini@genes.nig.ac.jp).

FEATURES	location/Qualifiers
source	1. .608

BASE COUNT	ORIGIN
153 a	135 c 142 g 177 t 1 others

Query Match	19.8%	Score 309.4	DB 9	Length 608
Best Local Similarity	69.2%	Pred. No. 5,1e-72		
Matches 421	Conservative 0	Mismatches 187	Indels 0	Gaps 0

358 accgcctggtggaataatcgattactctgtcaaatctctcgagatggtgcacataacga 417

Db	1	ACCGTCGGCGGCATTCGTGTGACGACGATACCGTGGCCCTCCCGCGGATTAATTGGCTGTT	60
Oy	418	ttggaatccctatgcttgatgaagaagaagatcgctttacacgtaagcttgagcgtgt	477
Db	61	CTTAAAGATCATCATTGATGAGGGGACATGCTAGTCATTTCTACACCGTCAGCGTGGCTTGT	120
Oy	478	ggcgttaatgynaacccataltgttcgagctbgagagtaaaagtlataatccagtcatt	537
Db	121	GACCTTGACGGCACACCCTGCTGTGTGGCACAGAAAGCAATGGGGTCATTCGGGGTATC	180
Oy	538	gaagtcacagtgtaagaagatccataaagctcttgtagtcaatgagatctagtgaaagaa	597
Db	181	AACGTGTGCCACCGAAGAGTTGTTTBACTTTTCTTGCCATGCGGATTCATTAATAATGAG	240
Oy	598	atcaggacacaaacctttaaactcleaacttgbatctactcgtatgaagaagtatcgtt	657
Db	241	ATAAGAACTCAACCATTAAGACCTTCATCTTTCATTTTGGCAAGCAAGAGAGAGCTTGT	300
Oy	658	cgcttgtagaatgttgaacccgggataagcttttgatcttgcgtaggcgcgaagtcatt	717
Db	301	AGGTTATGGAATGTCCATPACAGGGATGTGCATTGTGATTTTGCTGGAGAGAGAGGTAC	360
Oy	718	cgctatgaagcttcaagctgtgatttcaatccgcctcgtatattcaacgctttgtagtctgt	777
Db	361	CGTATATGAAGTATGAGTGTGATCTCCACCCTTGATATCTACCGAATGCCAGTTGT	420
Oy	778	ggtatgagacaccactataaataatgttcaatgaagaagctttgtagcgtatgaagaag	837
Db	421	GGCATGGATTAATACGTTTAAATCTGGTCAATGAAGAAGATTTTGGCCATACCTGGAGAA	480
Oy	838	tcaatcacatgtagctgtagatccatcacaaatccccaanaatctgtccaaatccctgta	897
Db	481	TGCTTTACATGAGCTGACCTTCCATCAAAATTTCCACAAAAATATGTCCAATTTCCACTT	540
Oy	898	tttaacgcttccatccaataaataatgtatagatgttbaacgcttgcttgggtatcttacc	957
Db	541	ATGACTTCGCTGGTGCATTTCTAATCATGTGTGACTGATTAAGGTGGCTGGTACTTTCATC	600
Oy	958	ccctcaaa 965	
Db	601	CTGTGAA 608	

RESULT	15
BGI32016	
LOCUS	BGI32016
DEFINITION	BGI32016 tomato crown gall Lycopersicon esculentum EST 31-JAN-2001
ACCESSION	ES1464908 tomato crown gall Lycopersicon esculentum cDNA clone
VERSION	CIO6Ekl1 5' sequence, mRNA sequence.
KEYWORDS	BGI32016 BGI32016 GI:12632204
	EST.

ORGANISM    *Lycopersicon esculentum*  
 Eukaryota: Viridiplantae: Streptophyta: Tracheophyta;

REFERENCE AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 566) van der Hoeven, R., Sun, H., Cho, J., Uterback, T., Hansen, C., Ronning, C. and Tanksley, S.	Generation of ESTs from tomato crown gall tissue	Unpublished (2001)	
	Contact: CUGI		

Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>

```

FEATURES
  source      location/Qualifiers
1. 566       /organism="Lycopersicon esculentum"
          /cultivar="TA96"
          /db_xref="taxon:4081"

```

```

/clone="C706E11"
/clone_lib="tomato crown gall"
/tissue_type="crown gall"
/dev_stage="crown galls from full-grown plants (8 wks old
)"
/lab_host="SOLR"
/note="vector: pBluescript SK(-); Site_1: EcoRI, Site_2:
XhoI; Four wk old greenhouse plants were stab inoculated
on stem with Agrobacterium tumefaciens C58 (Dr. T. J. Burr
Correll U.). Galls were allowed to develop for another 4
wks, when gall tissue was frozen in liquid nitrogen."
BASE COUNT      152 a      100 c      145 g      109 t
ORIGIN

```

[illegible]

Search completed: June 24, 2002, 23:23:07  
Job time: 3241 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 25, 2002, 01:13:56 ; Search time 88.61 Seconds

(without alignments)  
720.405 Million cell updates/sec

Title: US-09-812-283-4

Perfect score: 1973  
Sequence: 1 MSKITIGNESIVGSLRPSNR.....STLACCEDEGTWRMDVITR 369

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_19:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.rvirus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1973	100.0	369	10	Q9XF44
2	1565	79.3	294	10	Q9L747
3	727.5	36.9	441	4	000149
4	727	36.8	427	4	09UNY7
5	727	36.8	441	11	0921E6
6	727	36.8	517	4	075530
7	727	36.8	535	11	P97462
8	718	36.4	462	5	Q9VKD5
9	717	36.3	426	13	Q90YU5
10	715	36.2	425	5	024338
11	708	35.9	425	5	026458
12	703	35.6	428	5	016023
13	690	35.0	412	5	016021
14	689.5	34.9	437	5	016022
15	370	18.8	190	5	Q95SM6
16	300	15.2	459	5	016187

17	300	15.2	536	5	Q9GXS1	Q9GXS1 caenorhabd
18	211.5	10.7	665	5	Q9GNM6	Q9GNM6 caenorhabd
19	200	10.1	430	5	Q9B54	Q9B54 heterodera
20	197.5	10.0	913	10	Q9CAAO	Q9CAAO arabidopsis
21	197.5	10.0	920	10	Q9SDS5	Q9SDS5 arabidopsis
22	194.5	9.9	911	10	Q9LW87	Q9LW87 arabidopsis
23	194.5	9.9	926	10	Q9C827	Q9C827 arabidopsis
24	193.5	9.8	252	11	Q922C7	Q922C7 mus musculu
25	188	9.5	860	5	Q9VJ20	Q9VJ20 drosophila
26	187.5	9.5	421	4	Q9UBH9	Q9UBH9 homo sapien
27	187.5	9.5	421	4	Q9Y617	Q9Y617 homo sapien
28	186	9.4	421	11	054927	054927 mus musculu
29	185.5	9.4	569	11	Q92159	Q92159 mus musculu
30	183.5	9.3	421	4	Q9UNM6	Q9UNM6 homo sapien
31	183.5	9.3	1326	5	Q9VZF4	Q9VZF4 drosophila
32	182	9.2	377	10	Q9SW94	Q9SW94 pisum sativ
33	181.5	9.2	510	5	044382	044382 drosophila
34	181.5	9.2	510	5	Q9VDE3	Q9VDE3 drosophila
35	180.5	9.1	553	4	Q9NMX6	Q9NMX6 homo sapien
36	180.5	9.1	561	4	Q96R12	Q96R12 homo sapien
37	180.5	9.1	589	4	Q96LE0	Q96LE0 homo sapien
38	180.5	9.1	627	4	Q96A16	Q96A16 homo sapien
39	180.5	9.1	707	4	Q969H0	Q969H0 homo sapien
40	178.5	9.0	569	11	Q9QU15	Q9QU15 mus musculu
41	177	9.0	655	4	Q9BVA0	Q9BVA0 homo sapien
42	177	9.0	657	11	Q9CWM2	Q9CWM2 mus musculu
43	176.5	8.9	335	5	Q9BKG5	Q9BKG5 aedes aegypt
44	176.5	8.9	335	5	Q9BMG4	Q9BMG4 aedes aegypt
45	176.5	8.9	563	11	Q923H0	Q923H0 mus musculu

## ALIGNMENTS

RESULT 1  
Q9XF44 PRELIMINARY: PRT: 369 AA.  
AC Q9XF44;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE FERTILIZATION-INDEPENDENT ENDOSPERM PROTEIN.  
GN FILE  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. LANDSBERG ERRECTA;  
RX MEDLINE=99172055; PubMed=10072400;  
RA Ohad N., Yadegari R., Margossian L., Hannon M., Michaeli D.,  
RA Harada J.J., Goldberg R.B., Fischer R.L.;  
RT "Mutations in FIE, a WD polycarb group gene, allow endosperm  
development without fertilization.";  
RT Plant Cell 11:407-416(1999).  
RL -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
CC EMBL: AF129516; AAD23584.1; -;  
DR InterPro: IPR001680; WD40.  
DR Pfam: PR00400; WD40; 7.  
DR PRINTS: PR00320; GPROTEINRPT.  
DR SMART: SM00320; WD40; 3.  
DR PROSITE: PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
DR PROSITE: PS50082; WD\_REPEATS\_2; 2.  
DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
KM Repeat: WD repeat.  
SQ SEQUENCE 369 AA; 41257 MW; EDF0989C02B4FDDA CRC64;

Query Match 100.0%; Score 1973; DB 10; Length 369;  
Best local Similarity 100.0%; Pred. No. 1.7e-176;  
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	MSKITLGNESIVGSLTPENKKSYKVTNRIOGKRPLYAVFNEFLDAREPDEVATAGNRI	60
Db	1	MSKITLGNESIVGSLTPENKKSYKVTNRIOGKRPLYAVFNEFLDAREPDEVATAGNRI	60
OY	61	TLVNLGDAISALQSYADEDEKEESFYTVSMACGVNGNPIYAAAGVKIIIRYIDVNSETI	120
Db	61	TLVNLGDAISALQSYADEDEKEESFYTVSMACGVNGNPIYAAAGVKIIIRYIDVNSETI	120
OY	121	HKSLVGHSDSVNEIRTOPLKPOLVTTASKDESIVRLMNETGICILIFAGAGHREVLISV	180
Db	121	HKSLVGHSDSVNEIRTOPLKPOLVTTASKDESIVRLMNETGICILIFAGAGHREVLISV	180
OY	181	DEHPSDIYRFASCGMDTITIKITSMKEFPTYVEKSFPTWDDPSKFPETKVOEPTFASIIHT	240
Db	181	DEHPSDIYRFASCGMDTITIKITSMKEFPTYVEKSFPTWDDPSKFPETKVOEPTFASIIHT	240
OY	241	NYVDGNRMFGDFILISKSVDNELLMEPOLKENSPEGASDVLLARYVPMDCIWTFKFSCD	300
Db	241	NYVDGNRMFGDFILISKSVDNELLMEPOLKENSPEGASDVLLARYVPMDCIWTFKFSCD	300
OY	301	LHLSSVAIGNOEGKRYVNDLKSCEPVLLTKLSHNSKSVIRQTASVSGSTLLACCDEGT	360
Db	301	LHLSSVAIGNOEGKRYVNDLKSCEPVLLTKLSHNSKSVIRQTASVSGSTLLACCDEGT	360
OY	361	IWRMDVITK 369	
Db	361	IWRMDVITK 369	
RESULT	2		
O9LT47			
ID	O9LT47	PRELIMINARY;	PRT; 294 AA.
AC	O9LT47		
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)		
DE	FERTILIZATION-INDEPENDENT-ENDOSPERM PROTEIN-LIKE.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=COLUMBIA;		
RA	Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;		
RL	Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=COLUMBIA;		
RX	MEDLINE=20277480; PubMed=10819329;		
RA	Nakamura Y.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence		
RT	features of the regions of 4,504,864 bp covered by sixty P1 and TAC		
RT	clones."		
RL	DNA Res. 7:131-135(2000).		
CC	-I- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).		
DR	EMBL; AB025629; BAB02481.1; -		
DR	InterPro; IPR001680; WD40.		
DR	Pfam; PF00400; WD40; 6.		
DR	PRINTS; PR00320; GPROTEINBRPT.		
DR	SMART; SM00320; WD40; 3.		
DR	PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.		
DR	PROSITE; PS00682; WD_REPEATS_2; 2.		
DR	PROSITE; PS0294; WD_REPEATS_REGION; 1.		
KW	Repeat; WD repeat.		
SO	SEQUENCE 294 AA; 33024 MW; A9A2A1DC83A8B4CD CRC64;		
Query Match	79.3%;	Score 1565;	DB 10; Length 294;
Best Local Similarity	99.7%;	Pred. No. 2.2e-138;	
Matches 289; Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;

QY	80	EDKESFETVSMAGVGNPNPYAAGVKGKJIIIRVDVNSEFTHSKSLVGHGSVNEIRPPL	139
Db	5	EQKESFETVSMAGVGNPNPYAAGVKGKJIIIRVDVNSEFTHSKSLVGHGSVNEIRPPL	64
QY	140	KPOLYITASKDESRLNWNVEGTICILIFAGAGGHRVYLSVDFPSDIYFPAASGMOTI	199
Db	65	KPOLYITASKDESRLNWNVEGTICILIFAGAGGHRVYLSVDFPSDIYFPAASGMOTI	124
QY	200	KIMSMKEFWTVESKFTWTDPSKFEPTKFOVPYFETASIHYNVDCNRMGDFILSKSYD	259
Db	125	KIMSMKEFWTVESKFTWTDPSKFEPTKFOVPYFETASIHYNVDCNRMGDFILSKSYD	184
QY	260	NEILLMEPOLKENSPGBASDYLIRTPVPMCDIMFIFESCDLHLSVAIGNQESKYVMD	319
Db	185	NEILLMEPOLKENSPGBASDYLIRTPVPMCDIMFIFESCDLHLSVAIGNQESKYVMD	244
QY	320	LKSCPPVILITLSHNQSKSVIRQTAMSDGSTILLACCEDEGTIMRWMDIYTK	369
Db	245	LKSCPPVILITLSHNQSKSVIRQTAMSDGSTILLACCEDEGTIMRWMDIYTK	294

Query	18 SNKK---	SYKVTNRIOEG-KKPLVAVVFNEIDARFDD--VFPTAGGNRTLLVNCIGDGA	71
Matches	153;	Conservative 74; Mismatches 115;	Indels 29; Gaps 14;
Best Local Similarity	41.2%;	Pred. No. 1.3e-59;	
Score	727.5;	DB 4;	Length 441;
Sequence	441 AA;	50155 MW;	2EAE5BFEFEA56B0 CRC64;



```

Db 74 SSKCKYSEFKVNSLKEDHNOPLFGVQFMW-HSKEGDPLVFATVGSNRYLYECHSGEEL 132
Qy 72 SALQSTADDEKDESEFTYVSACGVN-GNPYYAAGVKGIIIRYIDVNSETIHKSLVGHGDS 130
Db 133 RLQSGVYDADADENFYTCAMTYDSNTSHPLAVAGSGRIIRINPTMOCIKHYVGHGNA 192
Qy 131 VNEIRTOPAKPOLVITASDESRLNVEGICILIFAGAGHREVLSDVHPSDIY--DLGEG 188
Db 193 INELKFNPDPLILSVSDHAKRLMINIOTDPLVAIFGVEGHRDEVLSADY---DLGEG 249
Qy 189 RASCGMDTTIKIW--SMKEFTYVEKSFPTWDDPSK---FPTKVOFPVF-TASIHYN 241
Db 250 KIMSCGMDHSLKIMRINSKRMMAIKESYD--NPNKTRNPFISQKIHPPDFSTRDIHRN 307
Qy 242 YVDCNRWEDFLSKSVNDEILMEP-QLKEN----SPEGASDVLYRPVPCMDIWFILK 296
Db 308 YVDCVWMLDGLILSKSCENAIYCMKPGKMKDDIDKIKPSESNTTILGRDYSCDDIWMYR 367
Qy 297 FSCDHLSSVAIGNOEGKYVWDLSKCP--VLITKLSHNSKSVYROTAMSVDSSTILA 354
Db 368 FSDMFQOKMLALGNQYKLYWDLVEDEPHKAKCTTLTHHKCGAIIROTSPFSKSSILIA 427
Qy 355 CCEGDTIWRMD 365
Db 428 VCDASITWRMD 438

```

```

RESULT 4
QUNY7 PRELIMINARY; PRT; 427 AA.
AC 09UNY7:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE WAIT-1.
GN WAIT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98438520; PubMed=9765275;
RA Riezler M., Bitner M., Kolanus W., Schuster A., Holzman B.;
RT "The human WD repeat protein WAIT-1 specifically interacts with the
RT cytoplasmic tails of beta7-integrins."
RL J. Biol. Chem. 273:27459-27466(1998).
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AF078933; AAC68675.1;
DR InterPro: IPR004048; RCCL1.
DR InterPro: IPR001680; WD40.
DR PIRam: PF00400; WD40; 5.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00626; RCCL1_2; UNKNOWN_1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE: PS00682; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 427 AA; 48779 MW; F2A4FE5A09CF9ACD CRC64;

```

Query Match 36.8%; Score 727; DB 4; Length 427;  
 Best Local Similarity 41.3%; Pred. No. 1.3e-59;  
 Matches 151; Conservative 70; Mismatches 119; Indels 26; Gaps 12;

```

Qy 20 KSKYKYNRIQEG-KKPLAAYVNFIDARFDD--VFVTAGNRIITLYNCIGDAISALOS 76
Db 65 KTSFKCVNSLKEDHNOPLFGVQFMW-HSKEGDPLVFATVGSNRYLYECHSGEIRLLOS 123
Qy 77 YADEKDESEFTYVSACGVN-GNPYYAAGVKGIIIRYIDVNSETIHKSLVGHGDSYNEIR 135
Db 124 YVDADADENFYTCAMTYDSNTSHPLAVAGSGRIIRINPTMOCIKHYVGHGNAINEK 183

```

```

Qy 136 TOPLPOLVITASKDESRLNVEGICILIFAGAGHREVLSDVHPSDIY--RFASC 193
Db 184 FPRDPNLLSLVSKDHALRLMINIOTDPLVAIFGVEGHRDEVLSADY---DLGEGIMSC 240
Qy 194 GMDTTIKIW--SMKEFTYVEKSFPTWDDPSK---FPTKVOFPVF-TASIHYNVDON 246
Db 241 GMDHSLKIMRINSKRMMAIKESYD--NPNKTRNPFISQKIHPPDFSTRDIHRNVDCY 298
Qy 247 RWFGDPLILSKSVNDEILMEPOLKEN----SPEGASDVLYRPVPCMDIWFILKSCDL 301
Db 299 RWFGLDILSKSCENAIYCMKPGKMEDDIDKIKPSESNTTILGRDYSCDDIWMYRSMDF 358
Qy 302 HLSVAIGNOEGKYVWDLSKCP--VLITKLSHNSKSVYROTAMSVDSSTILACEDG 359
Db 359 WQKMLALGNQYKLYWDLVEDEPHKAKCTTLTHHKCGAIIROTSPFSKSSILIAVCDDA 418
Qy 360 TIWRMD 365
Db 419 SIWRMD 424

```

```

RESULT 5
QUNY7 PRELIMINARY; PRT; 441 AA.
AC 0921E6:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SIMILAR TO EMBRYONIC ECTODERM DEVELOPMENT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC012966; AAH12966.1;
SQ SEQUENCE 441 AA; 50197 MW; D2E0A5BA27C0499A CRC64;

```

Query Match 36.8%; Score 727; DB 11; Length 441;  
 Best Local Similarity 41.3%; Pred. No. 1.4e-59;  
 Matches 151; Conservative 70; Mismatches 119; Indels 26; Gaps 12;

```

Qy 20 KSKYKYNRIQEG-KKPLAAYVNFIDARFDD--VFVTAGNRIITLYNCIGDAISALOS 76
Db 79 KTSFKCVNSLKEDHNOPLFGVQFMW-HSKEGDPLVFATVGSNRYLYECHSGEIRLLOS 137
Qy 77 YADEKDESEFTYVSACGVN-GNPYYAAGVKGIIIRYIDVNSETIHKSLVGHGDSYNEIR 135
Db 138 YVDADADENFYTCAMTYDSNTSHPLAVAGSGRIIRINPTMOCIKHYVGHGNAINEK 197
Qy 136 TOPLPOLVITASKDESRLNVEGICILIFAGAGHREVLSDVHPSDIY--RFASC 193
Db 198 FPRDPNLLSLVSKDHALRLMINIOTDPLVAIFGVEGHRDEVLSADY---DLGEGIMSC 254
Qy 194 GMDTTIKIW--SMKEFTYVEKSFPTWDDPSK---FPTKVOFPVF-TASIHYNVDON 246
Db 255 GMDHSLKIMRINSKRMMAIKESYD--NPNKTRNPFISQKIHPPDFSTRDIHRNVDCY 312
Qy 247 RWFGDPLILSKSVNDEILMEPOLKEN----SPEGASDVLYRPVPCMDIWFILKSCDL 301
Db 313 RWFGLDILSKSCENAIYCMKPGKMEDDIDKIKPSESNTTILGRDYSCDDIWMYRSMDF 372
Qy 302 HLSVAIGNOEGKYVWDLSKCP--VLITKLSHNSKSVYROTAMSVDSSTILACEDG 359
Db 373 WQKMLALGNQYKLYWDLVEDEPHKAKCTTLTHHKCGAIIROTSPFSKSSILIAVCDDA 432
Qy 360 TIWRMD 365
Db 433 SIWRMD 438

```

```

Oy      360  TIMEWD 365
          :|||||
Db      509  SIMRWD 514

RESULT      7
P97462
AC      P97462;      PRELIMINARY;      PRT;      535 AA.
DT      01-MAY-1997 (TIMBLrel. 03, Created)
DT      01-JAN-1998 (TIMBLrel. 05, Last sequence update)
DT      01-DEC-2001 (TIMBLrel. 19, Last annotation update)
DE      EMBRYONIC ECTODERM DEVELOPMENT (EMBRYONIC ECTODERM DEVELOPMENT
DE      PROTEIN).
GN      EED OR EED.
OS      MUS MUSCULUS (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.
OX      NCBI_TaxID=10990;
      [1]
      SEQUENCE FROM N.A.
RA      Denisenko O., Bomszlyk K.;
RL      Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
      [2]
      SEQUENCE OF 95-535 FROM N.A.
RP      STRAIN=SWISS WEBSTER;
RC      MEDLINE=96353973; PubMed=8717038;
RX      Schumacher A., Faust C., Magnuson T.;
RA      "Role of leptin in the neuroendocrine response to fasting.";
RT      Nature 382:250-253(1996).
CC      -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CL      EMBL: 097675; AAC53302.1; -.
DR      EMBL: 078103; AAB38319.1; -.
DR      MGD: MGI:95286; Bed.
DR      InterPro: IPR000408; RCCL.
DR      InterPro: IPR001680; WD40.
DR      Pfam: PF00400; WD40. 5.
DR      PRINTS: PR00320; GPROTEINBPT.
DR      SMART: SM00320; WD40. 3.
DR      PROSITE: PS00626; RCCL_2; UNKNOWN_1.
DR      PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR      PROSITE: PS00682; WD_REPEATS_2; 2.
DR      PROSITE: PS0294; WD_REPEATS_REGION; 1.
KW      Repeat; WD repeat.
SQ      SEQUENCE      535 AA.      59997 MW.      7FF1216AB6BE497A CRC64;

```

	Query Match	36.8%	Score 727;	DB 11;	Length 535;
	Best Local Similarity	41.3%;	Pred. No. 1.9e-59;		
	Matches 151; Conservative		70; Mismatches 119;	Indels 26;	Gaps 12.
Qy	20 KKSRYKTVRIIOGG-KKPLVAVFNEFLDAHFPP--VFVTAGGRITLYLNLGDAISALQS	76			
Dd	173 KTSFKCVNLSLKEDHNPRLPFGQFNW-HSMGEQDPLVFATGSRNRVLTECHSGOGEIRLLS	231			
Qy	77 YADEDEKESEFYIVSWACGVN-GNPYVAAGCGKIIRVIDVNSETIHKSLVGGDSDNER	135			
Dd	232 YVDADDEFNYFCAMTYDDNTSHPLAVALGSGIIIRINIPITMOCTIKHHYGGNALNEUK	291			
Qy	136 TQPLRQLVLTASKDSEVALMVNETGICILFAGAGRHREYLVSDVPHSPDIY--RFASC	193			
Dd	292 FHRDRNLLLSKDALRLMNIQTDTLVLAITIGVGHGHDVLSADY---DLIGERIMSC	348			
Qy	194 GMDTITIKIW-SMKEFWTVERSEFTWDPSK---EPTRVOQFPVE-TASIHTRYVCQN	246			
Dd	349 GMDSHLAKLRINSKRMMNAIKESGYD--NPKNTRNPFISQIKHEFPDFSTRDIRHNRYDCV	406			
Qy	247 RMFGDFILSKVDNELMLPEOLKEN-----SGEGEASDYLLRXYPMDDIWFIRKSCDL	301			
Dd	407 RMLGDELILSKSCENAIIVCWKPKGMEDIDIKIRPSESNVTJILGEDYSODCDIMWRSSMPF	466			
Qy	302 HLLSVAILNGEGKVUYWLDKSCP--VLITTKISHNQSKSVIRQTAMSYDGSTILLACEDG	359			



```

OY 77 VADEDEKESFYTVMAGVYV-NPNVYAGVGIIVIDVNSSTIHKSLVGHGDSVNEIR 135
DB 123 YVADADENFTYCAWTDSTSHPLAVAGSRGIIRIINPTIQCICKHYGHONAIINELK 182
OY 136 TOPLAKQOLVITASKDSVRLMNVETGICILIFAGAGHREVLVSDFPHSPDIY--FPASC 193
DB 183 FHRDRNLLLSVSKDHRLMNIOTDTLVAIFGVGSHREVLVSADY--DLGCKIMSC 239
OY 194 GMDTTIKRISMK--EFWTYVEKSFVTWDDPSK----FPRTYVOFPVY-TPASITNTVDCN 246
DB 240 GMDHSLKIMRINLRKAKTAKESYD--NPNKNRPVSOQKVPFPESTDIRHNTVDCV 297
OY 247 RWFDFILSKSVNEILLWEP-----OLKENSPEGASDVLLRYPPVPCDIWFKISCDL 301
DB 298 RMLGDLILSKSCENALYCMKPGKMEDDIEKIKASESNVTLIGRFDSQCDIWMYRSMDF 357
OY 302 HLLSVAIGNOEGVYVWDLSKSCP--VLITKLSHNSKSVYIROTAMVDSGTLIACEDG 359
DB 358 WOKMLALGNVGYKLYWDLVLEKDPHAKCTTLYPKCASAIROTSFRSDSVLIAVCDSD 417
OY 360 TIRWMD 365
DB 418 TIRWMD 423

RESULT 10
O24338 PRELIMINARY; PRT: 425 AA.
AC O24338;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ESC PROTEIN.
DE ESC OR G514941.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sulton G.G., Wotman J.R., Richards S., Ashburner M., Henderson S.N.,
RA George R.A., Lewis S.E., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Ball J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegan C.,
RA Jaitai M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle V.J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shie B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

```

RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96016202; PubMed=7556071;
RA Gujjar T., Frei E., Spicer C., Baumgartner S., White R.A., Noll M.;
RT "The Polycomb-group gene, extra sex combs, encodes a nuclear member of
RT the WD-40 repeat family.";
RL EMBO J. 14:4296-4306(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazer R.G.,
RA Butenoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Hummasti S.R., Kara K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nixon K., Pacle V.J.M., Park S., Pfeiffer B., Poon L., Punch E.,
RA Sequera A., Sethi H., Solt E., Svirska R.R., Twomey B., Wan K.H.,
RA Weinburg T., Zhang R., Zieran L.L., Rubin G.M.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AE003634; AAF53124.1; -.
DR EMBL: LA1867; AAA86427.1; -.
DR EMBL: AC006240; -. NOT_ANNOTATED_CDS.
DR TRANSFAC: T02156;
DR FlyBase: FBgn0000588; esc.
DR InterPro: IPR000408; RCCL.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 5.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE: PS50082; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ
SEQUENCE 425 AA; 47987 MW; 511C305E5D86727 CRC64;

Query Match 36.2%; Score 715; DB 5; Length 425;
Best Local Similarity 39.3%; Pred. No. 1.8e-58;
Matches 143; Conservative 76; Mismatches 129; Indels 16; Gaps 10;

OY 17 PSNKSIVYTRKIDEGK-KPLIYAVVFNLDARF-DVYVTAGNRITLNYCLDGSATSL 74
DB 56 PKSRAAKYVDYHVENHGANIFGAFTLLGKDEPOVFATGSKRVYVECPRGGMQL 115
OY 75 QSVADDEKESFYTVMAGVYV-NPNVYAGVGIIVIDVNSSTIHKSLVGHGDSVNE 133
DB 116 HCYADPDEDFYTCAMSYDLDKTSPLLAAGYGVIRIVDQENAVNGVYIGHGOALNE 175
OY 134 IRTQPLKQOLVITASKDSVRLMNVETGICILIFAGAGHREVLVSDFPHSPDIYRASC 193
DB 176 LKPHRKQQLLLSSKDHRLMNIOTDTLVAIFGVGSHREVLVSADY--DLGCKIMSC 234
OY 194 GMDTTIKRISMK--EFWTYVEKSFVTWDDPSK--KPPRTYVOFPVY-TPASITNTVDCN 248
DB 235 GMDHSLKIMRINLRKAKTAKESYD--NPNKNRPVSOQKVPFPESTDIRHNTVDCV 294
OY 247 RWFDFILSKSVNEILLWEP--OLKENSPEGASDVLLRYPPVPCDIWFKISCDL 303
DB 298 RMLGDLILSKSCENALYCMKPGKMEDDIEKIKASESNVTLIGRFDSQCDIWMYRSMDF 357
OY 302 HLLSVAIGNOEGVYVWDLSKSCP--VLITKLSHNSKSVYIROTAMVDSGTLIACEDGT 361
DB 358 WOKMLALGNVGYKLYWDLVLEKDPHAKCTTLYPKCASAIROTSFRSDSVLIAVCDSD 417

```

QY	362	WRWD	365
		:	
Db	415	WRWN	418

RESULT	11			
026458				
ID	026458	PRELIMINARY;	PRT:	425 AA.
AC	026458;			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	ESC.			
GN	ESC.			
OS	Drosophila virilis (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7244;			
RN	[1]			
RP	SEQUENCE FROM N.A..			
RX	MEDLINE=96081366; PubMed=8541211;			
RA	Sathe S.S., Harle P.J.;			
RT	"The extra sex combs protein is highly conserved between Drosophila			
RT	virilis and Drosophila melanogaster.";			
RL	Mech. Dev. 52:225-232(1995).			
CC	-1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).			
CC	EMBL; S80985; AAB35874.2; -.			
DR	EMBL; S80983; AAB35874.2; JOINED.			
DR	EMBL; S80983; AAB35874.2; JOINED.			
DR	EMBL; S80984; AAB35874.2; JOINED.			
DR	FLYBase; FBgn0013917; Df1Yesc.			
DR	InterPro; IPR000408; RCCL.			
DR	InterPro; IPR001680; WD40.			
DR	Pfam; PF00400; WD40; 5.			
DR	PRINTS; PR00320; GPROTEINBRPT.			
DR	SMART; SM00330; WD40_3.			
DR	PROSITE; PS00626; RCCL_2; UNKNOWN_1.			
DR	PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.			
DR	PROSITE; PS50082; WD_REPEATS_2; 2.			
DR	PROSITE; PS50294; WD_REPEATS_REGION; 1.			
KW	Repeat; WD repeat..			
SO	SEQUENCE	425 AA;	48042 MW;	EB5B0A0D5F5A93B8 CRC64;

[illegible]

QY	362	WRWD	365
		111:	
Db	415	WRWN	418

RESULT	12			
ID	016023	PRELIMINARY;	PRT;	428 AA.
AC	016023;			
DT	01-JAN-1998	(TREMBLrel. 05, Created)		
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	EXTRA SEX COMBS.			
GN	ESC.			
OS	Musca domestica (House fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Muscoidea; Muscidae; Musca.			
OX	NCHI_TaxID=7370;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE-96001595; Pubmed-9343430;			
RT	Ng J., Li R., Morgan K., Simon J.;			
RT	"Evolutionary conservation and predicted structure of the Drosophila			
RT	extra sex combs repressor protein."			
RL	Mol. Cell. Biol. 17:6663-6672(1997).			
CC	-1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).			
DR	EMBL; AF003605; AAC05333.1; -			
DR	InterPro; IPR000408; RCCL.			
DR	InterPro; IPR001680; WD40.			
DR	Pfam; PF00400; WD40; 6.			
DR	SMART; SMO0320; WD40; 3.			
DR	PROSITE; PS00626; RCCL_2; UNKNOWN_1.			
DR	PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.			
DR	PROSITE; PS00882; WD_REPEATS_2; 2			
DR	PROSITE; PS50294; WD_REPEATS_REGION; 1.			
FW	Repeat; WD repeat.			
SO	SEQUENCE 428 AA; 48796 MW; 541D09D63D1A86C2 CRC64;			

Query Match 35.6%; Score 703; DB 5; Length 428;  
Best Local Similarity 38.6%; Pred. No. 2.4e-57;  
Matches 156; Conservative 67; Mismatches 129; Indels 52; Gaps 13;

```

RESULT 13
ID 016021 PRELIMINARY: PRT: 412 AA.
AC 016021;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE EXTRA SEX COMBS.
GN ESC.
OS Junonia coenia (Peacock butterfly) (Precis coenia).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrystia;
OC Papilionoidea; Nymphalidae; Nymphalinae; Junonia.
OX NCBI_TaxID=39708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98001595; PubMed=9343430;
RA Ng J., Li R., Morgan K., Simon J.;
RT "Evolutionary conservation and predicted structure of the Drosophila
RT extra sex combs repressor protein."
RL Mol. Cell. Biol. 17:6663-6672(1997).
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AF003603; AAC05331.1; -.
DR InterPro: IPR001005; Myb_DNA_bind.
DR InterPro: IPR000408; RCCL.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 5.
DR PRINTS: PR00320; GPROTEINBRPT.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
DR PROSITE: PS00678; RCCL_2; UNKNOWN_1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE: PS50082; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 412 AA; 46757 MW; 10D47281BDFE2BC6 CRC64;

Query Match 35.0%; Score 690; DB 5; Length 412;
Best Local Similarity 38.0%; Pred. No. 3,7e-56;
Matches 150; Conservative 71; Mismatches 134; Indels 40; Gaps 12;

OY 5 TLGNESTVSTIPSNK-----SYKVTNRI-----QEGKKPLAAVFN-FLD 45
DB 17 TSNTDNTSRSEPTNTRVKKRRRKKKVKYQKKPKFNCSAKEDHGQPLFGCOFNHLG 76
OY 46 ARFDVFTAGNRTITVNCIGDAISALQSYADEDEKESEFYVSWACGVNGN--PYVAAG 104
DB 77 EGEPLVFAVGSNRYVSECPESGFKLQCYADPDVDENYTCAMSYDEETNLPFLAVA 136
OY 105 GVKGIIRVADNSETIHKSLVGHGDSVNEIRTOPKLPOLVITASKDESVRILMNVETGICI 164
DB 137 GSRGIIRVFHATQCTCIKHVGHGHAINEVFKHPDPMILLASAKDHALRLMINISDVCII 196
OY 165 LIFPAGAGHRLEVLSDVPHPSDI--YRPASGGMOTTIKI-----SMKEFTYVKSFTW 217
DB 197 AIFGVEGHRDEVLADP---DLGERITMSGMHSKILMRLDKSME--AIKQSYNF 250
OY 218 TDDPSKFTKFWO--FPVF--TASITNTVDCNRMGDFILSKSVNELLMEPOLKENS- 273
DB 251 NPHRLRPFNSIKHEFPFSTRDHRNRYVDCVWMGDLILSKSCNAILTCMKPGLEDETE 310
OY 274 --PGEASDVLRLRYFVPMCDIWFIFKSCDLHLSSVAIGNOEGKYVMDLKS--CPVLLTK 330
DB 311 LRPQGNSTYIYHRPDKCEIMFIRFAVDYSQRYALGNQCKTWMLGSAAGSRVSO 370
OY 331 LSHNOSKSVIRQTAMSVDSGTTILACCEGDTIWRMD 365
DB 371 LVHPRCAVAARQVTLRSNGKYLITCCDDGTIWRMD 405

```

```

ID 016022 PRELIMINARY: PRT: 437 AA.
AC 016022;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE EXTRA SEX COMBS.
GN ESC.
OS Schistocerca americana (American grasshopper).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC Acridoidea; Acrididae; Acrididae; Schistocerca.
OX NCBI_TaxID=7009;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98001595; PubMed=9343430;
RA Ng J., Li R., Morgan K., Simon J.;
RT "Evolutionary conservation and predicted structure of the Drosophila
RT extra sex combs repressor protein."
RL Mol. Cell. Biol. 17:6663-6672(1997).
CC -1- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AF003604; AAC05332.1; -.
DR InterPro: IPR000408; RCCL.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 4.
DR PRINTS: PR00320; GPROTEINBRPT.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00678; RCCL_2; UNKNOWN_1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
DR PROSITE: PS50082; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 437 AA; 49655 MW; 7F7C227A8D3E59E CRC64;

Query Match 34.9%; Score 689.5; DB 5; Length 437;
Best Local Similarity 38.6%; Pred. No. 4,5e-56;
Matches 151; Conservative 66; Mismatches 131; Indels 43; Gaps 12;

OY 14 SLTPSN-----KSKYKTNRIQEGK-KELAAVFN-FLDARFPDVF 52
DB 51 SEPTNRRGRKRRGRKKIKQKINNMKLOYKFSYVKEDEHGQPLFGAOFNHLKAGQPLIF 110
OY 53 VTAGNRTITVNCIGDAISALQSYADEDEKESEFYVSWAC-GVNGNRYVAAGVKGIIR 111
DB 111 AAVSNRRTYVECPGSGSIRKILQCYADPDVDENYTCAMSYDEESGKPLAVAGSRGIIR 170
OY 112 VIDVNSETIHKSLVGHGDSVNEIRTOPKLPOLVITASKDESVRILMNVETGICILPAGAG 171
DB 171 IFSPATLSCIRHYIGHGHAINEVFKHPDPMILLASAKDHALRLMINIKTDVCAIFGVE 230
OY 172 GHRREVLSVDHPHSDIY--RFASGGMOTTIKI-----SMKEFTYVKSFTWDDPSKF 224
DB 231 GHRREVLSADP---DLGERITMSGMHSKILMRLDKDSME--AVANSILFNSARSILR 284
OY 225 PTKFWO--FPVF--TASITNTVDCNRMGDFILSKSVNELLMEP--QLKENSPEGGA 278
DB 285 PDSLSKEHFPFSTRDHRNRYVDCVWMGDLILSKSCNAILTCMKPGLEKELKTNDT 344
OY 279 SDVLRYPVPMCDIWFIFKSCDLHLSSVAIGNOEGKYVMDLKSPP--VLITLSHNOS 336
DB 345 VTIIHREFYRCEIWFVFAFADFQKILALGNQCKTFVMDLVDSPPSQRCTALTTPRC 404
OY 337 KSVIRQTAMSVDSGTTILACCEGDTIWRMDVI 367
DB 405 VAAIRQTSLSRDSVYLCCDDGTIWRMDRI 435

```

RESULT 14  
016022

DT 01-DEC-2001 (Tremblrel. 19, last annotation update)  
DE SD02661P.  
GN CG5202.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.,  
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF060456; AAL25495.1; -  
SQ SEQUENCE 190 AA; 22015 MW; 5EBE1A3B99D06E14 CRC64;

Query Match 18.8%; Score 370; DB 5; Length 190;

Best Local Similarity 41.2%; Pred. No. 1.2e-26;  
Matches 77; Conservative 33; Mismatches 63; Indels 14; Gaps 6;

OY 192 SCGMDFTIRIW-----SMKEFWTVVEKSFPTWDDPSKFPKFPVQFPVF-TASIHFNVDG 245  
DB 2 SSGMDHSLKMRDKRDIKEA-IELSSGFSFNKNTGPFPIKEHFPDFSTROIHRNVDG 60  
OY 246 NRWEGDFILSKSVNDELMEP-QLKEN---SPGEGASDVILRYVPVPCDIWFIKFS 300  
DB 61 VWFVGFDFVFSKSCENSIVCKPGLSESWHEIKPQESATVLIHFDYKMCIEIWFVRFAFN 120  
OY 301 LHSLSVAIGNOEKGVVWMLKSCPVL--ITKLSHNSKSVIRQTAMSVDSITLACED 358  
DB 121 AMOKIILALGNQLGTFVWELDCNDPNLTKCSQLVHPKSNSTIROTSFSKDGSIIVCVCDD 180  
OY 359 GTIWRMD 365  
DB 181 STIWRMD 187

Search completed: June 25, 2002, 01:24:47  
Job time: 651 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 25, 2002, 00:14:26 ; Search time 67.93 Seconds

(Without alignments)  
132.681 Million cell updates/sec

Title: US-09-812-283-4

Perfect score: 1973  
Sequence: 1 MSKITLGNSIVSLTPSNK.....STLLACCEGCTIKRMVDYTK 369

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1973	100.0	369	4	US-09-177-249-4
2	202.5	10.3	704	1	US-08-190-802A-62
3	202.5	10.3	704	1	US-08-188-582-5
4	202.5	10.3	704	1	US-08-646-715-5
5	202.5	10.3	704	2	US-08-308-818-3
6	202.5	10.3	704	4	US-08-477-346-62
7	202.5	10.3	704	4	US-08-473-089-62
8	188	9.5	798	1	US-08-190-802A-64
9	188	9.5	798	1	US-08-190-802A-68
10	188	9.5	798	2	US-08-308-818-2
11	188	9.5	798	4	US-08-477-346-64
12	188	9.5	798	4	US-08-477-346-68
13	188	9.5	798	4	US-08-473-089-64
14	188	9.5	798	4	US-08-473-089-68
15	176.5	8.9	517	1	US-08-190-802A-30
16	176.5	8.9	517	4	US-08-477-346-30
17	176.5	8.9	517	4	US-08-473-089-30
18	175.5	8.9	517	4	US-08-473-089-30
19	170.5	8.6	514	4	US-08-190-802A-66
20	170.5	8.6	514	4	US-08-477-346-66
21	170.5	8.6	514	4	US-08-473-089-66
22	168	8.5	409	2	US-08-283-917-3
23	168	8.5	409	2	US-08-961-716-3
24	168	8.5	410	2	US-08-283-917-9
25	168	8.5	410	2	US-08-961-716-9
26	167.5	8.5	704	1	US-08-188-582-18
27	167.5	8.5	704	1	US-08-646-715-18

28	165.5	8.4	409	1	US-08-190-802A-51	Sequence 51, Appl
29	165.5	8.4	409	4	US-08-477-346-51	Sequence 51, Appl
30	165.5	8.4	409	4	US-08-473-089-51	Sequence 51, Appl
31	163	8.3	713	1	US-08-190-802A-63	Sequence 63, Appl
32	163	8.3	713	4	US-08-477-346-63	Sequence 63, Appl
33	163	8.3	713	4	US-08-473-089-63	Sequence 63, Appl
34	162.5	8.2	906	1	US-08-190-802A-31	Sequence 31, Appl
35	162.5	8.2	906	4	US-08-477-346-31	Sequence 31, Appl
36	162.5	8.2	906	4	US-08-473-089-31	Sequence 31, Appl
37	157.5	8.0	375	4	US-09-063-743-1	Sequence 1, Appl
38	153.5	7.8	317	1	US-08-190-802A-27	Sequence 27, Appl
39	153.5	7.8	317	1	US-08-190-802A-41	Sequence 41, Appl
40	153.5	7.8	317	1	US-08-190-802A-47	Sequence 47, Appl
41	153.5	7.8	317	4	US-08-477-346-27	Sequence 27, Appl
42	153.5	7.8	317	4	US-08-477-346-41	Sequence 41, Appl
43	153.5	7.8	317	4	US-08-473-089-47	Sequence 47, Appl
44	153.5	7.8	317	4	US-08-473-089-27	Sequence 27, Appl
45	153.5	7.8	317	4	US-08-473-089-41	Sequence 41, Appl

#### ALIGNMENTS

RESULT 1									
US-09-177-249-4									
Sequence 4, Application US/09177249									
Patent No. 6229064									
GENERAL INFORMATION:									
APPLICANT: Fischer, Robert L.									
APPLICANT: Ohad, Nir									
APPLICANT: Kiyose, Tomohiro									
APPLICANT: Yadegari, Ramit									
APPLICANT: Margossian, Linda									
APPLICANT: Harada, John									
APPLICANT: Goldberg, Robert B.									
APPLICANT: The Regents of the University of California									
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit									
TITLE OF INVENTION: Development in Plants									
FILE REFERENCE: 023070-08612005									
CURRENT FILING DATE: 1998-10-22									
EARLIER FILING DATE: 1998-05-01									
NUMBER OF SEQ ID NOS: 324									
SOFTWARE: PatentIn Ver. 2.0									
SEQ ID NO 4									
LENGTH: 369									
TYPE: PR									
ORGANISM: Arabidopsis sp.									
US-09-177-249-4									
Query Match									
Best Local Similarity 100.0%; Score 1973; DB 4; Length 369;									
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MSKITLGNSIVSLTPSNKSKSYVTNRIOEGKKPLVAVFNLARFDFVYTAGNRI	60						
DB	1	MSKITLGNSIVSLTPSNKSKSYVTNRIOEGKKPLVAVFNLARFDFVYTAGNRI	60						
QY	61	TLVNCIGDPAISLQSYADAEDEKESFYTSVACGVGNFYVAAAGVKGIIIRYDVNSRT	120						
DB	61	TLVNCIGDPAISLQSYADAEDEKESFYTSVACGVGNFYVAAAGVKGIIIRYDVNSRT	120						
QY	121	HSKIVGHGSVNEIRPQPLKPOLVITASKDESVRLNVEFGICILIFAGAGHRYEVL	180						
DB	121	HSKIVGHGSVNEIRPQPLKPOLVITASKDESVRLNVEFGICILIFAGAGHRYEVL	180						
QY	181	DEHPSDIYFPAAGMDTITIKINSKKEFTVYKESFTWDDPSKPKVQPFVFTASHT	240						
DB	181	DEHPSDIYFPAAGMDTITIKINSKKEFTVYKESFTWDDPSKPKVQPFVFTASHT	240						
QY	241	NYVDCNRWGGDFILSKSVNEILLMEPQLKENSFGEGASDVLIRYVPPACDIWTFISCD	300						



Db 526 ALRFVGG--HLSDVDCVQFHPNSNYATGSSDRTVRLM-----DNMTGQSVR-LMTGHKG 577  
QY 288 PMCDIWFIFKSCDHLHSSVAIGNQEGKYVMDLSCPPVLITKLSHNSQSVIRQTAMSV 347  
Db 578 SVSSLAFF--SACGRYLAS--GSVDHNIITLMDL-SNGSLVTTILRH--STVTTTTFSR 628  
QY 348 DGSFTIACCEDEGTIWRMD 365  
Db 629 DGTVLAAAGLDNNLTLMMD 646

## RESULT 4

US-08-646-715-5  
; Sequence 5, Application US/08646715  
; Patent No. 5637686

## GENERAL INFORMATION:

APPLICANT: Tjian, Robert  
APPLICANT: Comai, Lucio  
APPLICANT: Dylact, Brian D.  
APPLICANT: Hoey, Timothy  
APPLICANT: Ruppert, Siegfried  
APPLICANT: Tanese, Naoko  
APPLICANT: Wang, Edith  
APPLICANT: Weinzierl, Robert O.J.  
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESS: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/646,715  
FILING DATE: 09-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/188,582

FILING DATE: 28-JAN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A.

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 704 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-646-715-5

Query Match 10.3%; Score 202.5; DB 1: Length 704;

Best Local Similarity 25.2%; Pred. No. 2,2e-13;

Matches 65; Conservative 49; Mismatches 99; Indels 45; Gaps 13;

QY 110 IRVIDVNSETIHSLSVGHGDSVNEIRTPQLKPOLVITASDESVRIMNVETGICILIFAG 169

Db 432 VRMLDRSGEVTRSLMGHTGPYRCAPAP-EMMLLLSCSDSTIRLMSLTITWSCVVTYR- 489

QY 170 AGGHRREVLSVDHPSPDIYFAFGMDTTIKINSMEFTVYVKSFTWTDPSKFTPKFV 229

Db 490 --GHVYPMVDVREAPHGYY-FVSCSYDKTARLMA-----TDSNQ----- 525  
QY 230 QEPVFPASTHTNVDDCNRMF--GDPIKSVYDVEIILMEPQLKENSPEASVLLRYPV 287  
Db 526 ALRFVGG--HLSDVDCVQFHPNSNYATGSSDRTVRLM-----DNMTGQSVR-LMTGHKG 577  
QY 288 PMCDIWFIFKSCDHLHSSVAIGNQEGKYVMDLSCPPVLITKLSHNSQSVIRQTAMSV 347  
Db 578 SVSSLAFF--SACGRYLAS--GSVDHNIITLMDL-SNGSLVTTILRH--STVTTTTFSR 628  
QY 348 DGSFTIACCEDEGTIWRMD 365  
Db 629 DGTVLAAAGLDNNLTLMMD 646

## RESULT 5

US-08-308-818-3  
; Sequence 3, Application US/08308818  
; Patent No. 5847077

## GENERAL INFORMATION:

APPLICANT: Green, Michael R  
APPLICANT: Reese, Joseph C  
TITLE OF INVENTION: A No. 5847077el Fungal Multisubunit Protein  
TITLE OF INVENTION: Complex critical for Expression of Fungal Proteins  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESS: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: US  
ZIP: 10022

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/308,818  
FILING DATE: 19-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Ludwig, S. Peter

REGISTRATION NUMBER: 25,351

REFERENCE/DOCKET NUMBER: 0342/0A404

TELEPHONE: 212-52707700

TELEFAX: 212-753-6237

TELEX: 236687

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 704 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

FRAGMENT TYPE: N-terminal

ORGANISM: D. melanogaster

IMMEDIATE SOURCE:

CLONE: TAFIT-80

US-08-308-818-3

Query Match 10.3%; Score 202.5; DB 2: Length 704;

Best Local Similarity 25.2%; Pred. No. 2,2e-13;

Matches 65; Conservative 49; Mismatches 99; Indels 45; Gaps 13;

QY 110 IRVIDVNSETIHSLSVGHGDSVNEIRTPQLKPOLVITASDESVRIMNVETGICILIFAG 169

Db 432 VRMLDRSGEVTRSLMGHTGPYRCAPAP-EMMLLLSCSDSTIRLMSLTITWSCVVTYR- 489

QY 170 AGGHRREVLSVDHPSPDIYFAFGMDTTIKINSMEFTVYVKSFTWTDPSKFTPKFV 229

Db 490 --GHVPVMDVRFAPRGY-FVSCSYDKTARLMA-----IDSNO----- 525  
OY 230 QPVPFASHTNVYDCNRF--GDFILSKSVDEIILMEPOLKENSPEGASVILRPV 287  
Db 526 ALRVFVG--HLSVDCVQFHPNSNYATGSSDTRILM-----DNMTGQSVR-IMTGHKG 577  
OY 288 PMCDIFIKSCDLHLSVAIGNQEGKYVWDKSCPPVILTKLSHNSQSVIRQTAMSV 347  
Db 578 SVSLSAF--SACGRYLAS--GSVDHNIITIMDL-SNGSLVTLIRHT---STVTTITFSR 628  
OY 348 DGSTIIACCEDGTIRWD 365  
Db 629 DGTVLAAGLDNNLTLMWD 646

## RESULT 6

US-08-477-346-62  
; Sequence 62, Application US/08477346  
; Patent No. 6262023  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived peptides and uses  
; TITLE OF INVENTION: Theeof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morlison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,346  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/487,072  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2550-0025.20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-0763  
; TELEFAX: (202) 887-0763  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 704 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: TRANSCRIPTION FACTOR T1IF, Fig. 45  
; US-08-477-346-62

Query Match 10.3%, Score 202.5; DB 4; Length 704;  
Best Local Similarity 25.2%, Pred. No. 2.2e-13;  
Matches 65; Conservative 49; Mismatches 99; Indels 45; Gaps 13;  
OY 110 IRVIDVNSETHIKSLVYGDSDVNEIRQPLKPOLVITASKDSVRLMVNVTGICILIFAG 169  
Db 432 VMLDDRSGEYTRISLGHGTGPRYRCAPAP-EMNLLSCSESDSTIRLMSLITWSCVYTR- 489

OY 170 AGGHREYLVSDHPSPDIYRFASCGMDTTIKIWSMKEFWTYEKSFTWTDSPKFPTRFV 229  
Db 490 --GHVPVMDVRFAPRGY-FVSCSYDKTARLMA-----IDSNO----- 525  
OY 230 QPVPFASHTNVYDCNRF--GDFILSKSVDEIILMEPOLKENSPEGASVILRPV 287  
Db 526 ALRVFVG--HLSVDCVQFHPNSNYATGSSDTRILM-----DNMTGQSVR-IMTGHKG 577  
OY 288 PMCDIFIKSCDLHLSVAIGNQEGKYVWDKSCPPVILTKLSHNSQSVIRQTAMSV 347  
Db 578 SVSLSAF--SACGRYLAS--GSVDHNIITIMDL-SNGSLVTLIRHT---STVTTITFSR 628  
OY 348 DGSTIIACCEDGTIRWD 365  
Db 629 DGTVLAAGLDNNLTLMWD 646

## RESULT 7

US-08-473-089-62  
; Sequence 62, Application US/08473089  
; Patent No. 6342368  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived peptides and uses  
; TITLE OF INVENTION: Theeof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morlison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/473,089  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2550-0025.22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-0763  
; TELEFAX: (202) 887-0763  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 704 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: TRANSCRIPTION FACTOR T1IF, Fig. 45  
; US-08-473-089-62

Query Match 10.3%, Score 202.5; DB 4; Length 704;  
Best Local Similarity 25.2%, Pred. No. 2.2e-13;  
Matches 65; Conservative 49; Mismatches 99; Indels 45; Gaps 13;  
OY 110 IRVIDVNSETHIKSLVYGDSDVNEIRQPLKPOLVITASKDSVRLMVNVTGICILIFAG 169  
Db 432 VMLDDRSGEYTRISLGHGTGPRYRCAPAP-EMNLLSCSESDSTIRLMSLITWSCVYTR- 489  
OY 170 AGGHREYLVSDHPSPDIYRFASCGMDTTIKIWSMKEFWTYEKSFTWTDSPKFPTRFV 229

Db 490 --GHVYVMDVDFEAPHGY--FVSCSYDKTARLMA-----TDSNQ----- 525  
 QY 230 QPVPFLASHITNYVDCNRF--GDFILSKSVNELLMEPOLKENSPEGASVLLRYPV 287  
 Db 526 ALRVEFG--HLSVDVDFVOPHNSNYVATGSSDPTVRLW-----DNMTGGSVR-LMTGHKG 577  
 QY 288 PMCDIFIKFSCDLHLSSAIGNOEKGYVWDLKSCPVLITKLSINOSKSVIRQTAMSV 347  
 Db 578 SVSSIAF--SACGRYLA--GSVDHNIITLMDL-SNGSLVTILLRHT---SYVTITTFE 628  
 QY 348 DGSITLACCEDGTWRWD 365  
 Db 629 DGTVLAAGLDNNLTLM 646

## RESULT 8

US-08-190-802A-64  
 ; Sequence 64, Application US/08190802A  
 ; Patent No. 5519003  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mochly-Rosen, Daria  
 ; APPLICANT: Ron, Dorit  
 ; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
 ; TITLE OF INVENTION: Theoreof  
 ; NUMBER OF SEQUENCES: 265  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dehlinger & Associates  
 ; STREET: P.O. Box 60850  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94306-0850.  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/190,802A  
 ; FILING DATE: 01-FEB-1994  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fabian, Gary R.  
 ; REGISTRATION NUMBER: 33,875  
 ; REFERENCE/DOCKET NUMBER: 8600-0139  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 324-0880  
 ; TELEFAX: (415) 324-0960  
 ; INFORMATION FOR SEQ ID NO: 64:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 798 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; INDIVIDUAL ISOLATE: TUP1 HOMOLOG, Fig. 47  
 ; US-08-190-802A-64

Query Match 9.5%; Score 188; DB 1; Length 798;  
 Best Local Similarity 20.4%; Pred. No. 1.1e-11;

Matches 65; Conservative 64; Mismatches 110; Indels 80; Gaps 12;

QY 74 LOSYADEDEKEESFYTYSMAGVGNPNYVAGYKGIIRIVDVNSEITH----- 121  
 Db 459 MYTFQNTKMDSCLDSDDCRI-----AAAGFQDSYIKIWSLDGSSILNNPINALNNDKD 513  
 QY 122 ----KSLVGHGDSVNEIRTPQLKPOLVITASDESVRILNVEFGICILTFAGAGGHRYE 176  
 Db 514 EDPCTKTLVGHSGTVYSTSP--DNKYLLSGSDKTVRLMSMDTHTALVSK---GHNHP 569

QY 177 VLSVDFPSPDIYRFASCQMDFTIKIWSMKEFWTYVEKSFRTDPSKFPKFEVQPFETA 236  
 Db 570 VMVVSFSPGLGH-FATASHQOTARLMSCHI-----YLRIF 605  
 QY 237 SIHTNYVDCNRF--GDFILSKSVNELLMEPOLKENSPEGASVLLRYPVPMCDI-- 292  
 Db 606 AGHLNDVDCVSPHNCQYVFTGSSDKTCRMWDV-----STGDSVR-LFGLHTAPVISTAV 659  
 QY 293 ----WFIKSCDLHLSSAIGNOEKGYVWDLKSCPVLITKLSINOSKSVIRQTAMSV 347  
 Db 660 CPDGRW-----LSTGSDGIINWDTIGTKRL---KQMRGHGKNALYSLSYSK 704  
 QY 348 DGSITLACCEDGTWRWDV 366  
 Db 705 EGNVLISGGADHTVRYWDL 723

## RESULT 9

US-08-190-802A-68  
 ; Sequence 68, Application US/08190802A  
 ; Patent No. 5519003  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mochly-Rosen, Daria  
 ; APPLICANT: Ron, Dorit  
 ; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
 ; TITLE OF INVENTION: Theoreof  
 ; NUMBER OF SEQUENCES: 265  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dehlinger & Associates  
 ; STREET: P.O. Box 60850  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94306-0850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/190,802A  
 ; FILING DATE: 01-FEB-1994  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fabian, Gary R.  
 ; REGISTRATION NUMBER: 33,875  
 ; REFERENCE/DOCKET NUMBER: 8600-0139  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 324-0880  
 ; TELEFAX: (415) 324-0960  
 ; INFORMATION FOR SEQ ID NO: 68:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 798 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; INDIVIDUAL ISOLATE: yrb 1410 yeast, Fig. 51  
 ; US-08-190-802A-68

Query Match 9.5%; Score 188; DB 1; Length 798;  
 Best Local Similarity 20.4%; Pred. No. 1.1e-11;

Matches 65; Conservative 64; Mismatches 110; Indels 80; Gaps 12;

QY 74 LOSYADEDEKEESFYTYSMAGVGNPNYVAGYKGIIRIVDVNSEITH----- 121  
 Db 459 MYTFQNTKMDSCLDSDDCRI-----AAAGFQDSYIKIWSLDGSSILNNPINALNNDKD 513  
 QY 122 ----KSLVGHGDSVNEIRTPQLKPOLVITASDESVRILNVEFGICILTFAGAGGHRYE 176

Db 514 EDPCTCTLVGSGTGYSTSP-DNKYLLSGSEDKTVRLMSMDTFLALYSK---GHNHP 569  
QY 177 VLSVDHPSPDIYFPASCGMDTIKIKSMKEFWTVYKSFMTDDPSKFTKFPVQFPVETA 236  
Db 570 VMDVSPSPGLGHY-FATASHDQTLRLMSCDHI-----YPLRIF 605  
QY 237 SIHTNYVDCNRWF--GDFILSKSVNDEIILMEPOLKENSPEGASDYLRLYPVPMCDI-- 292  
Db 606 AGHLNDVDCVSPHNGCYVFTGSSDKTCRMWDV----STGDSVR-LFLGHTAPVLSIAV 659  
QY 293 -----WFIKSCDLHLSVAIGNQGVYVWDLKSCPPVILTLKSHNOSKSVIRQTAHSV 347  
Db 660 CPDGRW-----LSTGSEDIINWIDIGTGKRL--KOMRGHKNALYSLSYSK 704  
QY 348 DGSITLACCEGDTIWRWDV 366  
Db 705 EGNVLISGADHTYRWDL 723

RESULT 10  
US-08-308-818-2  
; Sequence 2, Application US/08308818  
; Patent No. 5847077  
; GENERAL INFORMATION:  
; APPLICANT: Green, Michael R  
; APPLICANT: Reese, Joseph C  
; TITLE OF INVENTION: A No. 5847077el Fungal Multisubunit Protein  
; NUMBER OF INVENTION: Complex Critical for Expression of Fungal Proteins  
; CORRESPONDENCE ADDRESS: 11  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: US  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/308,818  
; FILING DATE: 19-SEP-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, S. Peter  
; REGISTRATION NUMBER: 25,351  
; REFERENCE/DOCKET NUMBER: 0342/0A404  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-52707700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 798 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: S. cerevisiae  
; IMMEDIATE SOURCE:  
; CLONE: TAF-90  
; US-08-308-818-2

Query Match 9.5%; Score 188; DB 2; Length 798;  
Best Local Similarity 20.4%; Pred. No. 1,1e-11;  
Matches 65; Conservative 64; Mismatches 110; Indels 80; Gaps 12;  
QY 74 LQSYADEDEKESFYTVSMAGVNGNPNVYAGGVKGIINVIDVNSETH----- 121

Db 459 MTFFQNTKDMSCLDPSDDCRIT-----AAAGFQDSYIKIKSLDSSLNPNALNNNDK 513  
QY 122 -----KSLVGHGDSVNEIRTPOLPLVITASKDESVRILMNETGICILIFAGAGHRE 176  
Db 514 EDPCTCTLVGSGTGYSTSP-DNKYLLSGSEDKTVRLMSMDTFLALYSK---GHNHP 569  
QY 177 VLSVDHPSPDIYFPASCGMDTIKIKSMKEFWTVYKSFMTDDPSKFTKFPVQFPVETA 236  
Db 570 VMDVSPSPGLGHY-FATASHDQTLRLMSCDHI-----YPLRIF 605  
QY 237 SIHTNYVDCNRWF--GDFILSKSVNDEIILMEPOLKENSPEGASDYLRLYPVPMCDI-- 292  
Db 606 AGHLNDVDCVSPHNGCYVFTGSSDKTCRMWDV----STGDSVR-LFLGHTAPVLSIAV 659  
QY 293 -----WFIKSCDLHLSVAIGNQGVYVWDLKSCPPVILTLKSHNOSKSVIRQTAHSV 347  
Db 660 CPDGRW-----LSTGSEDIINWIDIGTGKRL--KOMRGHKNALYSLSYSK 704  
QY 348 DGSITLACCEGDTIWRWDV 366  
Db 705 EGNVLISGADHTYRWDL 723

RESULT 11  
US-08-477-346-64  
; Sequence 64, Application US/08477346  
; Patent No. 6262023  
; GENERAL INFORMATION:  
; APPLICANT: Ron, Dorit  
; APPLICANT: Mochly-Rosen, Daria  
; TITLE OF INVENTION: WD-40 - Derived peptides and Uses  
; NUMBER OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,346  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/487,072  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2550-0025.20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-0763  
; TELEFAX: (202) 887-1500  
; INFORMATION FOR SEQ ID NO: 64:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 798 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: TUP1 HOMOLOG, Fig. 47  
; US-08-477-346-64

Query Match 9.5%; Score 188; DB 4; Length 798;  
Best Local Similarity 20.4%; Pred. No. 1.1e-11;  
Matches 65; Conservative 64; Mismatches 110; Indels 80; Gaps 12;

QY 74 LOSYADEKKEEFTYVSMACVNGNPNYVAGVKGIIIRVIDNSETH-----121  
DB 459 MYTFONTNMDMCLDFSDDCRI-----AAAGFQDSYIKIWSLDGSSLNPNIALNNNDKD 513  
QY 122 -----KSLVGHDSVNEITOPKLPOLVITASKDESVRIMNVEGICILIFAGAGHRE 176  
DB 514 EDPCTKTLVGHSGTVSTFSF-DNKYLLSGSEDKTVRLMSDHTALVSYK---GHNP 569  
QY 177 VLSVDFHPSDIFRFASCGMDTTIKIWSMKEFTWYVKSFTWDDPSKFPTEVQFPVFA 236  
DB 570 VMDVFSFPLGH-FATASHDQARLMSCHI-----YPLRIE 605  
QY 237 SIHTNYDCNRMF--GDFLLSKSVNELLMEPQLKENSPEGASDVLRLYPVPMCDI-- 292  
DB 606 AGLHNDVDCVSEFHPNGCYFTGSSDKTCRMNDV-----STGDSVR-LFLGHTAPVISIAV 659  
QY 293 -----WFIKSCDLHLSVAIGNOEKVVYVWLKSCPPVLIKLHNSOKSVIROTAMSV 347  
DB 660 CPDGRW-----LSTGSEDGIINWMDIGTGKRL---KOMRGHKMAIYSLYSK 704  
QY 348 DGSLLACCEGDTIWRMDV 366  
DB 705 EGNVLISGADHTVRYWDL 723

## RESULT 12

US-08-477-346-68  
Sequence 68, Application US/08477346  
Patent No. 6262023

GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
APPLICANT: Ron, Dorit  
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
TITLE OF INVENTION: Theoreof  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 2000 Pennsylvania Avenue, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477.346  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487.072  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 798 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Yrb 1410 yeast, Fig. 51  
US-08-477-346-68

QY 74 LOSYADEKKEEFTYVSMACVNGNPNYVAGVKGIIIRVIDNSETH-----121  
DB 459 MYTFONTNMDMCLDFSDDCRI-----AAAGFQDSYIKIWSLDGSSLNPNIALNNNDKD 513  
QY 122 -----KSLVGHDSVNEITOPKLPOLVITASKDESVRIMNVEGICILIFAGAGHRE 176  
DB 514 EDPCTKTLVGHSGTVSTFSF-DNKYLLSGSEDKTVRLMSDHTALVSYK---GHNP 569  
QY 177 VLSVDFHPSDIFRFASCGMDTTIKIWSMKEFTWYVKSFTWDDPSKFPTEVQFPVFA 236  
DB 570 VMDVFSFPLGH-FATASHDQARLMSCHI-----YPLRIE 605  
QY 237 SIHTNYDCNRMF--GDFLLSKSVNELLMEPQLKENSPEGASDVLRLYPVPMCDI-- 292  
DB 606 AGLHNDVDCVSEFHPNGCYFTGSSDKTCRMNDV-----STGDSVR-LFLGHTAPVISIAV 659  
QY 293 -----WFIKSCDLHLSVAIGNOEKVVYVWLKSCPPVLIKLHNSOKSVIROTAMSV 347  
DB 660 CPDGRW-----LSTGSEDGIINWMDIGTGKRL---KOMRGHKMAIYSLYSK 704  
QY 348 DGSLLACCEGDTIWRMDV 366  
DB 705 EGNVLISGADHTVRYWDL 723

## RESULT 13

US-08-473-089-64  
Sequence 64, Application US/08473089  
Patent No. 6342368

GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
APPLICANT: Ron, Dorit  
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
TITLE OF INVENTION: Theoreof  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 2000 Pennsylvania Avenue, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473.089  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2550-0025.22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
INFORMATION FOR SEQ ID NO: 64:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 798 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

```

;      TYPE:  amino acid

```

SEQUENCE CHANGES TO:







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 25, 2002, 00:11:36 ; Search time 79.11 Seconds  
(without alignments)  
518.092 Million cell updates/sec

Title: US-09-812-283-4

Perfect score: 1973  
Sequence: 1 MSKITLGNSIVSLTPSNK.....STILACCEDGTIMRWVITK 369

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1973	100.0	369	AAV57037	Fertilisation-inde
2	1586	80.4	372	AAU00313	Fertilisation-inde
3	1576	79.9	370	AAU00315	Fertilisation-inde
4	1501	76.1	387	AAU00326	Fertilisation-inde
5	1402.5	71.1	379	AAU00309	Fertilisation-inde
6	1402.5	71.1	379	AAU00323	Fertilisation-inde
7	1394.5	70.7	379	AAU00321	Fertilisation-inde
8	1394	70.5	378	AAU00327	Fertilisation-inde
9	1391	70.5	375	AAU00325	Fertilisation-inde
10	1337.5	67.8	391	AAU00317	Fertilisation-inde
11	1219.5	61.8	461	AAU00311	Fertilisation-inde

12	1219.5	61.8	461	22	AAU00322	Fertilisation-inde
13	1215.5	61.6	363	22	AAU00339	Fertilisation-inde
14	1103	55.9	273	22	AAU00318	Fertilisation-inde
15	1045	53.0	430	22	AAU00320	Fertilisation-inde
16	720	36.5	162	22	AAU00342	Fertilisation-inde
17	718	36.4	462	22	ABB61196	Drosophila melanog
18	715	36.2	425	22	ABB69961	Drosophila melanog
19	708	35.9	158	22	AAU00335	Fertilisation-inde
20	662	33.6	152	22	AAU00336	Fertilisation-inde
21	648	32.8	141	22	AAU00341	Fertilisation-inde
22	590	29.9	145	22	AAU00330	Fertilisation-inde
23	590	29.9	155	22	AAU00312	Fertilisation-inde
24	507	25.7	129	22	AAU00334	Fertilisation-inde
25	498	25.2	125	22	AAU00314	Fertilisation-inde
26	486.5	24.7	119	22	AAU00332	Fertilisation-inde
27	462	23.4	121	22	AAU00316	Fertilisation-inde
28	456	23.1	124	22	AAU00333	Fertilisation-inde
29	422	21.4	119	22	AAU00324	Fertilisation-inde
30	365	18.5	98	22	AAU00340	Fertilisation-inde
31	335.5	17.0	116	22	AAU00329	Fertilisation-inde
32	309	15.7	111	22	AAU00343	Fertilisation-inde
33	264.5	13.4	84	22	AAU00328	Fertilisation-inde
34	220	11.2	59	22	AAU00331	Fertilisation-inde
35	216	10.9	59	22	AAU00338	Fertilisation-inde
36	202.5	10.3	704	16	AA85877	WP-40 domain-contg
37	202.5	10.3	704	17	AAW06079	Drosophila TATA-bi
38	202.5	10.3	704	18	AAW25027	TATA-binding prote
39	202.5	10.3	704	22	ABB63447	Drosophila melanog
40	198	10.0	55	22	AAU00337	Fertilisation-inde
41	198	10.0	55	22	AAU00319	Fertilisation-inde
42	197.5	10.0	704	15	AA856488	TATA-binding prote
43	194.5	9.9	542	21	AA856966	Human E3 ubiquitin
44	194.5	9.9	542	22	AAW79127	Human protein SEQ
45	194.5	9.9	542	22	AAW40208	Human polypeptide

#### ALIGNMENTS

RESULT 1  
AAV57037  
AAV57037 standard; Protein; 369 AA.  
XX  
AC AAV57037;  
XX  
DT 21-FEB-2000 (first entry)  
XX  
DE Fertilisation-independent endosperm 3 (FIE3) amino acid sequence.  
XX  
KW FIE3; fertilisation-independent endosperm; plant regulatory gene;  
KW reproductive development; polycarb gene; transgenic plant;  
KW plant morphology; cell cycle control; increase gene expression;  
KW vegetative growth.  
XX  
OS Arabidopsis sp.  
XX  
PN W09957247-A1.  
XX  
PD 11-NOV-1999.  
XX  
PF 03-MAY-1999; 99WO-US09676.  
XX  
PR 01-MAY-1998; 98US-0071838.  
XX  
PR 22-OCT-1998; 98US-0177249.  
XX  
PA (REGC ) UNTV CALIFORNIA.  
XX  
PI Fischer RL, Ohad N, Kiyosue T, Yadegari R, Margossian L, Harada J;  
PI Goldberg RB;  
XX  
DR WPI, 2000-023577/02.  
DR N-PSDB; AA237896.  
XX

PT FIE polynucleotide sequences, used to modulate endosperm development  
 XX  
 PS Claim 11; Page 35; 54pp; English.

CC This is the FIE3 (fertilisation-independent endosperm) amino acid  
 CC sequence. The FIE genes are a new group of plant regulatory genes. The  
 CC polypeptide encoded by the nucleotide sequence is involved in the  
 CC control of reproductive development. The FIE sequence shares homology  
 CC with a polycomb group gene from *Drosophila*. The invention relates to FIE1  
 CC and FIE3 polynucleotide sequences (AA37895-237896). The nucleotide  
 CC sequences can be used to create a transgenic plant containing an  
 CC expression cassette with an FIE nucleotide sequence operably linked to a  
 CC plant promoter. The FIE polynucleotide can be used to modulate endosperm  
 CC development in a plant. The nucleotide sequences can be used to increase  
 CC or enhance endogenous gene expression e.g. to enhance the expression of  
 CC polycomb genes which control cell cycling. This can be used to control  
 CC plant morphology. Enhanced expression can also be used to increase  
 CC vegetative growth.

XX  
 XX Sequence 369 AA:

Query Match 100.0%; Score 1973; DB 21; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-210;  
 Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKTTGNEISIVSLTPSNKSKYKVTNRIOGKKPLTAAVVFNDARFPVFTAGNR 60  
 DB 1 MSKTTGNEISIVSLTPSNKSKYKVTNRIOGKKPLTAAVVFNDARFPVFTAGNR 60  
 QY 61 TLVNCGLDGAISALQSYADEDESEFYTWSMAGVNGNPVAAAGVGKGIIRVIDVNSE 120  
 DB 61 TLVNCGLDGAISALQSYADEDESEFYTWSMAGVNGNPVAAAGVGKGIIRVIDVNSE 120  
 QY 121 HKSLVGHGDSVNEIRIOPKPLQVITASKDESVALMVEFGICLLIFAGAGHREYLSV 180  
 DB 121 HKSLVGHGDSVNEIRIOPKPLQVITASKDESVALMVEFGICLLIFAGAGHREYLSV 180  
 QY 181 DFRPSDIYRFASCGMDTITIKWSMKEFWTVESFTWTDDPSKPTKFPVFPFTASIR 240  
 DB 181 DFRPSDIYRFASCGMDTITIKWSMKEFWTVESFTWTDDPSKPTKFPVFPFTASIR 240  
 QY 241 NYVDCNRFEGDFILSKSVNDIELMEPOLKENSPGEGASDVLTARYPMCDIMEIKNSCD 300  
 DB 241 NYVDCNRFEGDFILSKSVNDIELMEPOLKENSPGEGASDVLTARYPMCDIMEIKNSCD 300  
 QY 301 LHSLSVAIGNOEKVYVWDKSCPPVLTIKLSHNSQSVIRQPMASVDGSRITLACCEGDT 360  
 DB 301 LHSLSVAIGNOEKVYVWDKSCPPVLTIKLSHNSQSVIRQPMASVDGSRITLACCEGDT 360  
 QY 361 IWRWDVITK 369  
 DB 361 IWRWDVITK 369

RESULT 2

AAU00313 standard; Protein; 372 AA.

XX  
 AC AAU00313:

XX 12-SEP-2001 (first entry)

DE Fertilisation-independent endosperm protein, eelc.pK003.e23.

KW Fertilisation-independent endosperm; plant reproduction; apomixis.

XX seed; pharmaceutical; nutraceutical; polymer; eelc.pK003.e23; corn.

OS Eucalyptus grandis.

XX  
 PN WO200116325-A2.  
 PD 08-MAR-2001.

XX 30-AUG-2000; 2000WO-US23735.  
 PF  
 XX 31-AUG-1999; 99US-0151575.  
 PR  
 XX

PA (DUPO) DU PONT DE NEMOURS & CO E. I.  
 PA (PION) PIONEER HI-BRED INT INC.

PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;  
 PI Simmons CR, Weng Z;

DR WPI: 2001-244407/25.  
 DR N-PSDB; AAS01160.

PT New plant fertilization independent endosperm protein for the  
 PT production of seed without fertilization is recombinantly produced -  
 XX  
 PS Claim 1; Page 45-47; 94pp; English.

CC The sequence represents the amino acid sequence of fertilisation-  
 CC independent endosperm protein, encoded by cDNA clone eelc.pK003.e23.  
 CC Fertilisation-independent endosperm proteins are plant reproduction  
 CC proteins necessary for apomixis, the formation of seeds without  
 CC fertilisation. Apomixis is especially useful to agriculture because it  
 CC eliminates the necessity of selfing plants to produce genetically  
 CC identical seed. Such seed is useful to produce seeds for human and animal  
 CC food and for commercial milling and extraction, including the production  
 CC of useful recombinant products in the endosperm e.g. pharmaceutical,  
 CC nutraceutical, industrial compounds and polymers. Embryoless seed  
 CC production by transgenic plants is less likely to cause ethical and  
 CC environmental concern over transgenic plant production as no gametes are  
 CC being produced to cross pollinate with other crops and the seeds cannot  
 CC germinate.

SO Sequence 372 AA:

Query Match 80.4%; Score 1586; DB 22; Length 372;  
 Best Local Similarity 75.5%; Pred. No. 2e-167;  
 Matches 278; Conservative 43; Mismatches 47; Indels 0; Gaps 0;

QY 1 MSKTTGNEISIVSLTPSNKSKYKVTNRIOGKKPLTAAVVFNDARFPVFTAGNR 60  
 DB 1 MSKTTGNEISIVSLTPSNKSKYKVTNRIOGKKPLTAAVVFNDARFPVFTAGNR 60  
 QY 61 TLVNCGLDGAISALQSYADEDESEFYTWSMAGVNGNPVAAAGVGKGIIRVIDVNSE 120  
 DB 61 TLVNCGLDGAISALQSYADEDESEFYTWSMAGVNGNPVAAAGVGKGIIRVIDVNSE 120  
 QY 121 HKSLVGHGDSVNEIRIOPKPLQVITASKDESVALMVEFGICLLIFAGAGHREYLSV 180  
 DB 121 HKSLVGHGDSVNEIRIOPKPLQVITASKDESVALMVEFGICLLIFAGAGHREYLSV 180  
 QY 181 DFRPSDIYRFASCGMDTITIKWSMKEFWTVESFTWTDDPSKPTKFPVFPFTASIR 240  
 DB 181 DFRPSDIYRFASCGMDTITIKWSMKEFWTVESFTWTDDPSKPTKFPVFPFTASIR 240  
 QY 241 NYVDCNRFEGDFILSKSVNDIELMEPOLKENSPGEGASDVLTARYPMCDIMEIKNSCD 300  
 DB 241 NYVDCNRFEGDFILSKSVNDIELMEPOLKENSPGEGASDVLTARYPMCDIMEIKNSCD 300  
 QY 301 LHSLSVAIGNOEKVYVWDKSCPPVLTIKLSHNSQSVIRQPMASVDGSRITLACCEGDT 360  
 DB 301 LHSLSVAIGNOEKVYVWDKSCPPVLTIKLSHNSQSVIRQPMASVDGSRITLACCEGDT 360  
 QY 361 IWRWDVITK 368  
 DB 361 IWRWDVITK 368

RESULT 3

AAU00315 standard; Protein; 370 AA.

[illegible][illegible]

```

Db      16  lqddpvwsglaesckkrejrvtnrlqegkrplvavlnfidszyfnvatvgsnrvtlygc 75
Oy      66  LGDGAISALOSTYADDEKKEESFTYVSMACGVGNCPVYAAAGVKGIIRVYDINSETIHKSLV 12
Db      76  ldegdlavlygsyadedknesfylvagacnvddtplvwaagllngvrlvdaagsekllhskfv 13
Oy      126 GHGDSVNEIRPOLPOLVITATSKDESRYLWVMEIGICILIFAGAGGHYEVLSVDHFHS 18
Db      136 gqgdsinevkaqllnpslvassakdesirlmahngicillfaga9ghnevlsvdfhps 19
Oy      186 DIYRPAESGMPTTIKIMSKKEFMPTVEKSFYWTDDPSKRPETKEFVOPVPEVTASIHFNYYDC 24
Db      196 dmyrltscgmsdtvtkwsmkeltwkyeksstvtldpskrptkfvqpytaashvlnyvc 25
Oy      246 NMFQDFILSKSVNDEILMEPOLKENSPEGASDVLRYPVPMCDIWEIKFSCDLHLSS 30
Db      256 ntwlgdfllsksvdnehlwepkveqprgkvadvllhkpypdcmiwlfikescdfhfnl 31
Oy      306 VAIGNQEGVYVWMDLKSCEPVLITKLISHQSSVIRQOTMASVDSGITLACCEDEGTWRMD 36
Db      316 vlvgnregklifweylqssppvllaakshpsqskspirlqtatstfdgstllscceqgtlwrwd 37
Oy      366 V 366
Db      376 v 376

```

RESULT 5  
 AAU00309  
 ID AAU00309 standard; Protein; 379 AA.  
 XX  
 AC AAU00309;  
 XX  
 DT 12-SEP-2001 (first entry)  
 XX  
 DE Fertilisation-independent endosperm protein, ccase-b.pk0026.g4.  
 XX  
 KW Fertilisation-independent endosperm; plant reproduction; apomixis;  
 XX seed; pharmaceutical; nutraceutical; polymer; ccase-b.pk0026.g4; corn.  
 XX  
 OS Zea mays.  
 XX  
 PN WO200116325-A2.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 30-AUG-2000; 2000MO-US23735.  
 XX  
 PR 31-AUG-1999; 99US-0151575.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;  
 XX Simmons CR, Weng Z;  
 DR MPI: 2001-24407/25.  
 XX N-PSDB; AAS01156.  
 PT  
 PT New plant fertilisation independent endosperm protein for the  
 XX production of seed without fertilization is recombinantly produced -  
 XX  
 PS Claim 1; Page 39-40; 94p; English.  
 XX  
 XX The sequence represents the amino acid sequence of fertilisation-  
 CC independent endosperm protein, encoded by cDNA clone ccase-b.pk0026.g4,  
 CC isolated from Zea mays. Fertilisation independent endosperm proteins are  
 CC plant reproduction proteins necessary for apomixis, the formation of  
 CC seeds without fertilisation. Apomixis is especially useful to agricultural  
 CC because it eliminates the necessity of selfing plants to produce  
 CC genetically identical seed. Such seed is useful to produce seeds for  
 CC human and animal food and for commercial milling and extraction,  
 CC including the production of useful recombinant products in the endosperm

CC e.g.pharmaceutical, nutraceutical, industrial compounds and polymers.  
CC Embryoless seed production by transgenic plants is less likely to cause  
CC ethical and environmental concern over transgenic plant production as no  
CC gametes are being produced to cross pollinate with other crops and the  
CC seeds cannot germinate.  
CC  
XX  
SQ Sequence 379 AA;

Query Match	71.1%	Score 1402.5	DB 22	Length 379
Best Local Similarity	63.48	Pred. No. 5.5e-147		
Matches 252; Conservative	48	Mismatches 62	Indels 1	Gaps 1

```

QY 6 LGNESIVSLPPSMKSKYKVTNRLOEGKKRPLVAVFNEFLAREFDVEVZAGNRITLYNC 65
Db 10 lgcceaaegslvpstkrfeykpcqkhtegktrlpaygfmfmdaryddvatcyggnrrvtlyrc 69
QY 66 LDCGASIALSGYADBEDKKEEFYFTVSMACG-VGNGPNVYAAGVGKGIIRVIDVNSETIHKSL 124
Db 70 lengsfalllggyvedekdesfytllswardhvdgspilvvaagsngilrtvlnccateklaxsf 129
QY 125 VGHGDSVNEIRTPOLKPOLVITAKSDSVRLANVETGICILIRAGAGGHRVLYLSDFHR 184
Db 130 vghgdsineirtprkpslilsaakdesvrlmwnhtgicililfagagghnevlsvdfhr 189
QY 165 SDIYFASCGMDTIKISMKEFFNTYVEKSTTWDDPSKPTKFPVORPPVFTASIIHNYYVD 244
Db 190 sdierfascgmndtvlksmkefvllyvdksyswcdlpsktrpkkyvgdfrpvliaavhsnyvd 249
QY 245 CNRRWGGDTLSKSVDNELLMEPOLKRENSPEGASADVLIRPAPMCDIMEIKRSCDLHLS 304
Db 250 ctnrwgdtflksyvdnevlmeprtkesppegsidllqkyrpeccilwflkiscdthfn 309
QY 305 SVAIGNOGKXYVWMDLKSPPVLTITKLSHNOSKSVIRQTAMSVDSGITTLACEDGTTIMRW 364
Db 310 qtaignuregkilywewgvsppvlliarlyngqckspirqtavsfsgstlllgagedgtltwv 369
QY 365 DVI 367
Db 370 dev 372

```

	RESULT	6
AAU00323	ID	AAU00323 standard; Protein: 379 AA.
XX	AC	AAU00323;
XX	DT	12-SEP-2001 (first entry)
XX	DE	Fertilisation-independent endosperm protein, p0107.cbca179r.
XX	KW	Fertilisation-independent endosperm; plant reproduction; apomixis
XX	KM	seed; pharmaceutical; nutraceutical; polymer; p0107.cbca179r.
OS	Zea mays.	
PN	WO200116325-A2.	
PD	08-MAR-2001.	
XX	30-AUG-2000; 2000MO-US23735.	
XX	31-AUG-1999; 99US-0151575.	
PA	(DUPO ) DU PONT DE NEMOURS & CO E I.	
PA	(PION-) PIONEER HI-BRED INT INC.	
PI	Butler KH, Danilevskaia O, Miao, G,	Morgante M, Sakai H;
WT:	2001-244407/25.	
N-PSDB:	AAS01170.	

XX New plant fertilization independent endosperm protein for the  
PT production of seed without fertilization is recombinantly produced -  
XX  
XX  
PS Claim 1; Page 63-64; 94pp; English.

XX The sequence represents the amino acid sequence of fertilisation-  
CC independent endosperm protein, encoded by cDNA clone p0107.cbca479r.  
CC Fertilisation-independent endosperm proteins are plant reproduction  
CC proteins necessary for apomixis, the formation of seeds without  
CC fertilisation. Apomixis is especially useful to agriculture because it  
CC eliminates the necessity of selfing plants to produce genetically  
CC identical seed. Such seed is useful to produce seeds for human and animal  
CC food and for commercial milling and extraction, including the production  
CC of useful recombinant products in the endosperm e.g. pharmaceutical,  
CC nutraceutical, industrial compounds and polymers. Embryolless seed  
CC production by transgenic plants is less likely to cause ethical and  
CC environmental concern over transgenic plant production as no gametes are  
CC being produced to cross pollinate with other crops and the seeds cannot  
CC germinate.  
XX  
SQ Sequence 379 AA:

Query Match 71.1%; Score 1402.5; DB 22; Length 379;  
Best Local Similarity 69.4%; Pred. No. 5.5e-147;  
Matches 252; Conservative 48; Mismatches 62; Indels 1; Gaps 1;

OY 6 LGNESIVGSLTPSNKSKSYVTNRIOEGKKPLVAVVNFIDARFEDVFTAGNRTLYNC 65  
DB 10 lgcEaaegslvpsrkreykpcgkhegkrlpYalgnfmdarydvfatvggnrvltc 69  
OY 66 LGDGAISALQSYADDEKESFYTVSWACG-VNGNPPYVAGVKGITRVIDVNSEITIKSL 124  
DB 70 lengsfallqayvdeekdesfYlswardhvdspllvaaagngllrvlncteklaksf 129  
OY 125 VGHGDSVNEIRPOLPLKPOLVITASKDESVRLNVEGICILIFAGAGGRIEVLSDVFNH 184  
DB 130 vghgdsineirtpkplksllsaskdesvrlwvnhgiclllfagaggrnevlsvdfhn 189  
OY 185 SDIYRFASCGMDTTIKMSKKEFWTVYVEKSFMTDPPSKFPKRVQFPVFTASIHNTYVD 244  
DB 190 sdierfascgmndtvkfwskewkfwlvvdskyswtldpskfpkyqfvpYllaavhnyvd 249  
OY 245 CNRMFGDFTLSKSVNELLMEPOLKENSPPGASDVLLRYVPMKDIWFIRFSCDLHLS 304  
DB 250 cctwlgdftllsksvndelwepkteqspgsgidllqkyppecdiwlfifscdfhn 309  
OY 305 SVAIGNOEKRVYVMDLKSPPVILITKLSHNSQSVYRQTAMSVDSSTILACCEGDTIWM 364  
DB 310 qiaignregklyvweqspvllarlynqgckspirqlavsfdsstlllgagedgtlwrw 369  
OY 365 DVI 367  
DB 370 dev 372

RESULT 7  
AAU00321  
ID AAU00321 standard; Protein; 379 AA.

XX AAU00321;  
XX  
XX 12-SEP-2001 (first entry)  
XX  
XX Fertilisation-independent endosperm protein, p0101.cgamg48r.  
XX  
XX Fertilisation-independent endosperm; plant reproduction; apomixis;  
XX seed; pharmaceutical; nutraceutical; polymer; p0101.cgamg48r.  
XX  
OS Zea mays.  
XX  
PN WO200116325-A2.

XX 08-MAR-2001.  
XX  
XX 30-AUG-2000; 2000WO-US23735.  
XX  
XX  
XX 31-AUG-1999; 99US-0151575.  
XX  
XX  
XX (DUPO ) DU PONT DE NEMOURS & CO E. I.  
XX (PION-) PIONEER HI-BRED INT INC.  
XX  
XX Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;  
XX Simmons CR, Weng Z;  
XX  
XX WPI: 2001-24407/25.  
XX  
XX N-PSDB: AAS01168.

PT New plant fertilization independent endosperm protein for the  
PT production of seed without fertilization is recombinantly produced -  
XX  
XX  
PS Claim 1; Page 58-60; 94pp; English.

CC The sequence represents the amino acid sequence of fertilisation-  
CC independent endosperm protein, encoded by cDNA clone p0101.cgamg48r.  
CC Fertilisation-independent endosperm proteins are plant reproduction  
CC proteins necessary for apomixis, the formation of seeds without  
CC fertilisation. Apomixis is especially useful to agriculture because it  
CC eliminates the necessity of selfing plants to produce genetically  
CC identical seed. Such seed is useful to produce seeds for human and animal  
CC food and for commercial milling and extraction, including the production  
CC of useful recombinant products in the endosperm e.g. pharmaceutical,  
CC nutraceutical, industrial compounds and polymers. Embryolless seed  
CC production by transgenic plants is less likely to cause ethical and  
CC environmental concern over transgenic plant production as no gametes are  
CC being produced to cross pollinate with other crops and the seeds cannot  
CC germinate.  
XX  
SQ Sequence 379 AA:

Query Match 70.7%; Score 1394.5; DB 22; Length 379;  
Best Local Similarity 68.9%; Pred. No. 4.3e-146;  
Matches 250; Conservative 48; Mismatches 64; Indels 1; Gaps 1;

OY 6 LGNESIVGSLTPSNKSKSYVTNRIOEGKKPLVAVVNFIDARFEDVFTAGNRTLYNC 65  
DB 10 lgcEaaegslvpsrkreykpcgkhegkrlpYalgnfmdarydvfatvggnrvltc 69  
OY 66 LGDGAISALQSYADDEKESFYTVSWACG-VNGNPPYVAGVKGITRVIDVNSEITIKSL 124  
DB 70 lengsfallqayvdeekdesfYlswardhvdspllvaaagngllrvlncteklaksf 129  
OY 125 VGHGDSVNEIRPOLPLKPOLVITASKDESVRLNVEGICILIFAGAGGRIEVLSDVFNH 184  
DB 130 vghgdsineirtpkplksllsaskdesvrlwvnhgiclllfagaggrnevlsvdfhn 189  
OY 185 SDIYRFASCGMDTTIKMSKKEFWTVYVEKSFMTDPPSKFPKRVQFPVFTASIHNTYVD 244  
DB 190 sdierfascgmndtvkfwskewkfwlvvdskyswtldpskfpkyqfvpYllaavhnyvd 249  
OY 245 CNRMFGDFTLSKSVNELLMEPOLKENSPPGASDVLLRYVPMKDIWFIRFSCDLHLS 304  
DB 250 cctwlgdftllsksvndelwepkteqspgsgidllqkyppecdiwlfifscdfhn 309  
OY 305 SVAIGNOEKRVYVMDLKSPPVILITKLSHNSQSVYRQTAMSVDSSTILACCEGDTIWM 364  
DB 310 qiaignregklyvweqspvllarlynqgckspirqlavsfdsstlllgagedgtlwrw 369  
OY 365 DVI 367  
DB 370 dev 372

RESULT 8

AAU00327  
 ID AAU00327 standard; Protein; 378 AA.  
 AC AAU00327;  
 DT 12-SEP-2001 (first entry)  
 DE Fertilisation-independent endosperm protein, ses2w.pk0015.b10.  
 XX Fertilisation-independent endosperm; plant reproduction; apomixis;  
 KM seed; pharmaceutical; nutraceutical; polymer; wheat; ses2w.pk0015.b10.  
 OS Triticum aestivum.  
 XX  
 PN MO200116325-A2.  
 PD 08-MAR-2001.  
 XX  
 PF 30-AUG-2000; 2000MO-US23735.  
 XX  
 PR 31-AUG-1999; 99US-0151575.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Butler KH, Danilevskaia O, Miao G, Morgante M, Sakai H;  
 PI Simmons CR, Weng Z;  
 DR WPI: 2001-244407/25.  
 DR N-PSDB; AAS01174.  
 PT New plant fertilization independent endosperm protein for the  
 PT production of seed without fertilization is recombinantly produced -  
 XX  
 PS Claim 1; Page 69-70; 94pp: English.  
 XX  
 CC The sequence represents the amino acid sequence of fertilisation-  
 CC independent endosperm protein, encoded by cDNA clone ses2w.pk0015.b10.  
 CC Fertilisation-independent endosperm proteins are plant reproduction  
 CC proteins necessary for apomixis, the formation of seeds without  
 CC fertilisation. Apomixis is especially useful to agriculture because it  
 CC eliminates the necessity of selfing plants to produce genetically  
 CC identical seed. Such seed is useful to produce seeds for human and animal  
 CC food and for commercial milling and extraction, including the production  
 CC of useful recombinant products in the endosperm e.g. pharmaceutical,  
 CC nutraceutical, industrial compounds and polymers. Embryoless seed  
 CC production by transgenic plants is less likely to cause ethical and  
 CC environmental concern over transgenic plant production as no gametes are  
 CC being produced to cross pollinate with other crops and the seeds cannot  
 CC germinate.  
 XX  
 SO Sequence 378 AA;

Query Match 70.7%; Score 1394; DB 22; Length 378;  
 Best Local Similarity 68.5%; Pred. NO. 4.8e-146;  
 Matches 248; Conservative 49; Mismatches 65; Indels 0; Gaps 0;

QY 6 LGNESIVSLPTPSKKSKVKNRQEGKKLYAVENFLDRFDEVTAGNRITLYNC 65  
 II : IIII : : : : : IIII : : : : : IIII : : : : : IIII : : : : :  
 DB 10 lgcnaavslaprsrrekyklsckhkgktrpyalgnldarydvatvgngnvtlyrg 69  
 QY 66 LGDGAISALOSYADPEDEKSEFYTWAGVGNPNVYAGVKGITRVIDVNSETIHKSLV 125  
 I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :  
 DB 70 lpdgnlavlgyidaddsgsfyltsacddlgdtrpllvagsgnavlrvincateklfkfsl 129  
 QY 126 GHGDSVNEIRTPQLKPOLVITASKDESVRLMNVETGICILIFAGAGGRREYLSDFRPS 185  
 IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :  
 DB 130 ghgdsneirtpqlkpslflfsskdesvrlwvhtgicillfagsggrnnevlsvdfhps 189  
 QY 186 DIYRASCAGMTTITINSMKEFTVVEKSFMTWDDPSKPPKRFVGFPEFASIHNNYDC 245  
 IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :  
 DB 190 diyrascgmdhtvklwsmkelpvveksftwclipskfpkfvqfplmtsvvshnydc 249

QY 246 NRMFGDILSKSVNDNEILLMEPOLKENSFGASDVLRLRPVPMCDIMFIRFSCDLHSS 305  
 II : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :  
 DB 250 trwlgdflfslsvdneiywepkiksgpgeisidvlkypdpdcdivfkfscdfhng 309  
 QY 306 VAIGNQEGKRVYWDLKSCPVLITRLSHNOSKSVIRQTAHMSYDGSSTIIACCDDGRTIWRMD 365  
 : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII : IIII :  
 DB 310 laignregklywewgaspvvlitrlsspqcmprqtavsfidgstllacgedgslytwd 369  
 QY 366 VI 367  
 :  
 DB 370 ev 371

RESULT 9  
 AAU00325  
 ID AAU00325 standard; Protein; 375 AA.  
 AC AAU00325;  
 DT 12-SEP-2001 (first entry)  
 DE Fertilisation-independent endosperm protein, p0120.cdebd48r.  
 XX  
 XX Fertilisation-independent endosperm; plant reproduction; apomixis;  
 KM seed; pharmaceutical; nutraceutical; polymer; rice; p0120.cdebd48r.  
 XX  
 OS Oryza sativa.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 279 /label= unknown  
 FT /note= "Encoded by gan"  
 XX  
 PN MO200116325-A2.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 30-AUG-2000; 2000MO-US23735.  
 XX  
 PR 31-AUG-1999; 99US-0151575.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Butler KH, Danilevskaia O, Miao G, Morgante M, Sakai H;  
 PI Simmons CR, Weng Z;  
 DR WPI: 2001-244407/25.  
 DR N-PSDB; AAS01172.  
 PT New plant fertilization independent endosperm protein for the  
 PT production of seed without fertilization is recombinantly produced -  
 XX  
 PS Claim 1; Page 65-67; 94pp: English.  
 XX  
 CC The sequence represents the amino acid sequence of fertilisation-  
 CC independent endosperm protein, encoded by cDNA clone p0120.cdebd48r.  
 CC Fertilisation-independent endosperm proteins are plant reproduction  
 CC proteins necessary for apomixis, the formation of seeds without  
 CC fertilisation. Apomixis is especially useful to agriculture because it  
 CC eliminates the necessity of selfing plants to produce genetically  
 CC identical seed. Such seed is useful to produce seeds for human and animal  
 CC food and for commercial milling and extraction, including the production  
 CC of useful recombinant products in the endosperm e.g. pharmaceutical,  
 CC nutraceutical, industrial compounds and polymers. Embryoless seed  
 CC production by transgenic plants is less likely to cause ethical and  
 CC environmental concern over transgenic plant production as no gametes are  
 CC being produced to cross pollinate with other crops and the seeds cannot  
 CC germinate.  
 XX  
 SO Sequence 375 AA;



Query Match 70.5%; Score 1391; DB 22; Length 375;  
 Best Local Similarity 69.2%; Pred. No. 1e-145;  
 Matches 245; Conservative 48; Mismatches 61; Indels 0; Gaps 0;

CC Fertilisation-independent endosperm proteins are plant reproduction  
 CC proteins necessary for apomixis, the formation of seeds without  
 CC fertilisation. Apomixis is especially useful to agriculture because it  
 CC eliminates the necessity of selfing plants to produce genetically  
 CC identical seed. Such seed is useful to produce seeds for human and animal  
 CC food and for commercial milling and extraction, including the production  
 CC of useful recombinant products in the endosperm e.g. pharmaceutical,  
 CC nutraceutical, industrial compounds and polymers. Embryoless seed  
 CC production by transgenic plants is less likely to cause ethical and  
 CC environmental concern over transgenic plant production as no gametes are  
 CC being produced to cross pollinate with other crops and the seeds cannot  
 CC germinate.

Sequence 391 AA:

Query Match 67.8%; Score 1337.5; DB 22; Length 391;  
 Best Local Similarity 70.7%; Pred. No. 9.9e-140;  
 Matches 239; Conservative 45; Mismatches 53; Indels 1; Gaps 1;

CC Fertilisation-independent endosperm protein, p0003.cgp29rb.  
 CC seed; pharmaceutical; nutraceutical; polymer; p0003.cgp29rb.  
 CC Zea mays.

Key Location/Qualifiers  
 FT Misc-difference 340 /note= "Encoded by gc"  
 XX WO200116325-A2.  
 XX PD 08-MAR-2001.  
 XX PF 30-AUG-2000; 2000WO-US23735.  
 XX PR 31-AUG-1999; 99US-0151575.  
 XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX PA (PION-) PIONEER HI-BRED INT INC.  
 XX PI Butler KH, DanilevsKaya O, Miao G, Morgante M, Sakai H;  
 XX PI Simmons CR, Weng Z;  
 XX DR WPI: 2001-244407/25.  
 XX DR N-PSDB: AAS01164.  
 XX PT New plant fertilization independent endosperm protein for the  
 XX production of seed without fertilization is recombinantly produced -  
 XX Claim 1, Page 52-53; 94pp; English.  
 XX The sequence represents the amino acid sequence of fertilisation-  
 CC independent endosperm protein, encoded by cDNA clone p0003.cgp29rb.

CC Fertilisation-independent endosperm proteins are plant reproduction  
 CC proteins necessary for apomixis, the formation of seeds without  
 CC fertilisation. Apomixis is especially useful to agriculture because it  
 CC eliminates the necessity of selfing plants to produce genetically  
 CC identical seed. Such seed is useful to produce seeds for human and animal  
 CC food and for commercial milling and extraction, including the production  
 CC of useful recombinant products in the endosperm e.g. pharmaceutical,  
 CC nutraceutical, industrial compounds and polymers. Embryoless seed  
 CC production by transgenic plants is less likely to cause ethical and  
 CC environmental concern over transgenic plant production as no gametes are  
 CC being produced to cross pollinate with other crops and the seeds cannot  
 CC germinate.

Sequence 391 AA:

Query Match 67.8%; Score 1337.5; DB 22; Length 391;  
 Best Local Similarity 70.7%; Pred. No. 9.9e-140;  
 Matches 239; Conservative 45; Mismatches 53; Indels 1; Gaps 1;

CC Fertilisation-independent endosperm protein, p0003.cgp29rb.  
 CC seed; pharmaceutical; nutraceutical; polymer; cen3npk0076.b8.  
 CC Zea mays.

Key Location/Qualifiers  
 FT Misc-difference 340 /note= "Encoded by gc"  
 XX WO200116325-A2.  
 XX PD 08-MAR-2001.  
 XX PF 30-AUG-2000; 2000WO-US23735.  
 XX PR 31-AUG-1999; 99US-0151575.  
 XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX PA (PION-) PIONEER HI-BRED INT INC.  
 XX PI Butler KH, DanilevsKaya O, Miao G, Morgante M, Sakai H;  
 XX PI Simmons CR, Weng Z;  
 XX DR WPI: 2001-244407/25.  
 XX DR N-PSDB: AAS01164.  
 XX PT New plant fertilization independent endosperm protein for the  
 XX production of seed without fertilization is recombinantly produced -  
 XX Claim 1, Page 52-53; 94pp; English.  
 XX The sequence represents the amino acid sequence of fertilisation-  
 CC independent endosperm protein, encoded by cDNA clone p0003.cgp29rb.

XX	WIPI: 2001-244407/25.
DR	N-PSDB: AAS01158.
XX	
PT	New plant fertilization independent endosperm protein for the
XX	production of seed without fertilization is recombinantly produced -
PS	Claim 1, Page 42-44; 94pp; English.
XX	
CC	The sequence represents the amino acid sequence of fertilisation-
CC	independent endosperm protein, encoded by cDNA clone cendp0076.18.
CC	Fertilisation-independent endosperm proteins are plant reproduction
CC	proteins necessary for apomixis, the formation of seeds without
CC	fertilisation. Apomixis is especially useful to agriculture because it
CC	eliminates the necessity of selfing plants to produce genetically
CC	identical seed. Such seed is useful to produce seeds for human and animal
CC	food and for commercial milling and extraction, including the production
CC	of useful recombinant products in the endosperm e.g. pharmaceutical,
CC	nutraceutical, industrial compounds and polymers. Embryotests seed
CC	production by transgenic plants is less likely to cause ethical and
CC	environmental concern over transgenic plant production as no gametes are
CC	being produced to cross pollinate with other crops and the seeds cannot
XX	germinate.
XX	
SO	Sequence 461 AA;

Query Match	61.88;	Score 1219.5;	DB 22;	Length 461;
Best Local Similarity	59.08;	Pred. No. 1.7e-126;		
Matches 214;	Conservative 63;	Mismatches 85;	Indels 1;	Gaps 1.

```

0Y      6 LGNESIVSLTPSPNNKKSvKYVNNRIQEOGCKPLTVAVFNFLARFPDVPVTAGGNITLTYNC 65
        | | | | | | | | | | : | : | | | | | | | | | | | | | | | | | | | |
Db     68 lgcvevegellvpsrtkreykpmskytvgnhpylaigfnfildmryydvafaiscmsvilyrc 127
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y     66 lEDGAISALQSvAEDBKDEESEFYTYSMAC- vCNGNPYYAAAGCVKCIIRYIDVNSETIHKSL 124
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    128 lengfglllqnyvedckdesfytlstldygdsapllvaagsnrlrvlnocateklaksl 187
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y    125 VGHGSVNEIRTPQLPOLVTTASKDSVRRLMNVETGICILIFGAGAGHREVLVSDFHP 184
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    188 yghgssilheirthaskpslllsasckdeslrilmwhbtglclivfagagghrndrldvsdfhp 247
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y    185 SDIYFPASCGMDTITIKISMKEFMFTYVKSFPTWDDPSKKPTKRVOPPVFFASHTNYVD 244
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    248 tevgfiascgmdntlvksmeketwlyekesyswtcgnhpskfptrniqfpvltaahsdgyd 307
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y    245 CARRWEFGDFlTSKYVDNELILMEPOLKENSPEGASDVLLRPVPMCDIMFKFSDDLHS 304
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    308 crrrwgdfllfsxvnavllwepkrpdckrrpgesavdrqkypvpkcslwfmkfcfdtsyn 367
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y    305 SVAIGNQBGRKYVWDLKSCPPYLITKLSHNOSKSVYRGTANMSVSGSTILACCEGGITWRM 364
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    368 gmaignkbeliyvwegvspyllrdlrlncegeckspqrctavsfsgstllgaadgaalwrw 427
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y    365 DVI 367
        | | |
Db    428 dev 430

```

RESULT	12
AAU00322	
ID	AAU00322 standard; Protein; 461 AA

AC AAU00322;

DT 12-SEP-2001 (first entry)

DE Fertilisation-independent endosperm protein, p0104.cabbn62r.

KW Fertilisation-independent endosperm; plant reproduction; apomixis;  
KW seed; pharmaceutical; nutraceutical; polymer; p0104.cabn62r.  
KW

XX

OS Zea mays.

PN WO200116325-A2.

PD 08-MAR-2001.

PF 30-AUG-2000; 2000WO-US23735.

PR 31-AUG-1999; 99US-0151575.

PA (DUPO ) DU PONT DE NEMOURS & CO E I .

XX	Danilevskaya O.	Miao G.	Morgante M.	Sakai H.
PI				
Butler KH				

PI Simmons CR, Weng Z;  
XY

DR WPI; 2001-24440//25.  
DR N-PSDB: AAS01169.

XX  
PT  
New plant fertilizati

XX production of seed without relling

PS Claim 1; page 60-62; 34pp; English.  
XX

The sequence represents the amino acid sequence of fertilisation-independent endosperm protein, encoded by cDNA clone p1014.cabn62r. Fertilisation-independent endosperm proteins are plant reproduction proteins necessary for apomixis, the formation of seeds without fertilisation. Apomixis is especially useful to agriculture because it eliminates the necessity of selling plants to produce genetically identical seed. Such seed is useful to produce seeds for human and animal food and for commercial milling and extraction, including the production of useful recombinant products in the endosperm e.g. pharmaceutical, nutraceutical, industrial compounds and polymers. Embryoleless seed production by transgenic plants is less likely to cause ethical and environmental concern over transgenic plant production as no gametes are being produced to cross pollinate with other crops and the seeds cannot germinate.

SQ Sequence 461 AA;

Query Match	61.8%	Score 1219.5	DB 22	Length 461
Best Local Similarity	59.0%	Pred. No. 1.7e-126		
Matches 214	Conservative	85	Indels 1	Gaps 1

	6	LONESIVSLPNNKSKYVNRNROEGKKPLAVFENFLARFDVVWAGRRITLYNC	65
QY		:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	68	Igcveevgellvpstkrrejkpmnskyvynghpyalgaifnildmryydvafalscscvilyrc	127
QY	66	LCDGALISALQSTADDEKDKESEFTVYSMACG-VGNCPYYAAGSVKGLIRVIDNVSETIHKSL	124
Dd	128	lenggfgllyqnvddckdesfyllswtldqvdsppllvaagsmclrtvlncaaceklajskl	187
QY	125	VGHGSVNEIRFOPKLPOLVITASKDESVRIMANNETGCILIFAGAGHRYEVLSDVFHP	184
Dd	188	vghgsjheirtnhakpsllsaakdesirilwnvhfgicllilfavagayrhdivlsavfnp	247
QY	195	Sltvpaasgcmmtktyskssfeefnyvrksptlmwmdndeskkpftkvdprrpfvtaaihtnyvd	244

Db 248 tevglfascgmantvklwsmkelwlyevensyswgnpskptnrdqfpvltaaavnsyvd 307

QY 245 CNRMFGDFILSKSVDNELLMEPOLKENSPEGASDVLLRYVPVMDIWFITKFSCDHL 304

Db 308 ctwlgdflksknavllwepkpdkrrpgsgsvdviqkypvpkcsllwllkiscalsn 307

305 SVAIGNOEGKVYVMDLKRSCPVLITKLSHNQSKSVIRQIAMSVDGSIILACCEGGIIMNM 304

Db 368 qmailgnkgeilyvwevqsspvlldrlcnqeckspilrtqtavsldgslldgaaduqalw 427

QY	365	DVI	367
	1	:	
438	439	439	439

Dib

RESULT 13  
AAU00339 standard; Protein; 363 AA.  
ID AAU00339 standard; Protein; 363 AA.  
AC AAU00339;  
XX  
XX  
DT 12-SEP-2001 (first entry)  
DE Fertilisation-independent endosperm protein, p0107.cbca179r.  
XX  
XX Fertilisation-independent endosperm; plant reproduction; apomixis;  
KW seed; pharmaceutical; nutraceutical; polymer; p0107.cbca179r. corn.  
XX  
OS Zea mays.  
XX  
XX WO200116325-A2.  
XX  
XX PD 08-MAR-2001.  
XX  
XX PF 30-AUG-2000; 2000WO-US23735.  
XX  
XX PR 31-AUG-1999; 99US-0151575.  
XX  
XX PA (DUPO ) DU PONT DE NEMOURS & CO E. I.  
XX  
XX PA (PION-) PIONEER HI-BRED INT INC.  
XX  
XX PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;  
XX  
XX PI Simmons CR, Weng Z;  
XX  
XX DR WPI, 2001-244407/25.  
XX  
XX PT New plant fertilization independent endosperm protein for the  
XX  
XX production of seed without fertilization is recombinantly produced -  
XX  
XX PS Claim 1; Page 84-85; 94pp; English.  
XX  
XX CC The sequence represents the amino acid sequence of fertilisation-  
XX  
XX independent endosperm protein, encoded by cDNA clone p0107.cbca179r.  
XX  
XX CC Fertilisation-independent endosperm proteins are plant reproduction  
XX  
XX proteins necessary for apomixis, the formation of seeds without  
XX  
XX fertilisation. Apomixis is especially useful to agriculture because it  
XX  
XX eliminates the necessity of selfing plants to produce genetically  
XX  
XX identical seed. Such seed is useful to produce seeds for human and animal  
XX  
XX food and for commercial milling and extraction, including the production  
XX  
XX of useful recombinant products in the endosperm e.g. pharmaceutical,  
XX  
XX nutraceutical, industrial compounds and polymers. Embryoleless seed  
XX  
XX production by transgenic plants is less likely to cause ethical and  
XX  
XX environmental concern over transgenic plant production as no gametes are  
XX  
XX being produced to cross pollinate with other crops and the seeds cannot  
XX  
XX germinate.  
XX  
XX CC Note: It is stated in the specification that this sequence is encoded by  
XX  
XX the DNA sequence shown in AAS01186, however this is not the case.  
XX  
XX SQ Sequence 363 AA;

Query Match 61.6%; Score 1215.5; DB 22; Length 363;  
Best Local Similarity 58.7%; Pred. NO. 3.3e-126;  
Matches 213; Conservative 63; Mismatches 86; Indels 1; Gaps 1;

QY 6 LGNESVSGSLRNSKSKSYVTRIDEGKRPVLYAVVFNFLDARFDVFTAGNRITLLNC 65  
DB 1 lgcveevgllypsrkreyknkyltgnphlylqifnldmryydvfaascnsvilyrc 60  
QY 66 LGDGAISALQSYADEDKESFTVSWAGC-VNGNPPVAAAGVKGITRIYDIVNSETHKSL 124  
DB 61 lenggflllgnvydeekdesfytiswtldqvdspllvaaagrrllrvlncatekidxsl 120  
QY 125 VGHGDSVNEIRTPQLKPOLVITASKDESVRLMNVETGICILFAGAGGRHYEVLVDVPH 184  
DB 121 vghggsiheltrhaskpsllisaskdesirlnwvhtclilvfagagrhndvlsvdfnp 180

QY 185 SDIYRFASCGMDTTIKWMSKEFWYVEKSPFTWDDPSKFPTRKFOVPVFTASIHNNVD 244  
DB 181 tevgifascgmhvtvkiwsketwlyekswtgnpskfpitrlngfpltaavnsdydv 240  
QY 245 CNRMFGDFTLKSVNDSEILLMEPOLKENSPPGASDYLRLVPMKIDIFIRFSCDLHS 304  
DB 241 ctrwlgdfllkskvknaevllwepkdkrrpgsgsvdlqkypkcsllwfmfscdfysn 300  
QY 305 SVAIGNQEGKRVYWDLKSCEPPVLITKLSHNSKSVIRQTAMSYDGTIIACCEBDGTIMRW 364  
DB 301 gmaignmkgeilywewgsspvllidrlcnqeckspirlqavsfdsstllggaaddgaalwrw 360  
QY 365 DVI 367  
DB 361 dev 363

RESULT 14  
AAU00318 standard; Protein; 273 AA.  
ID AAU00318 standard; Protein; 273 AA.  
AC AAU00318;  
XX  
XX  
DT 12-SEP-2001 (first entry)  
DE Fertilisation-independent endosperm protein, p0003.cgpf34rb.  
XX  
XX Fertilisation-independent endosperm; plant reproduction; apomixis;  
KW seed; pharmaceutical; nutraceutical; polymer; p0003.cgpf34rb.  
XX  
XX OS Zea mays.  
XX  
XX PN WO200116325-A2.  
XX  
XX PD 08-MAR-2001.  
XX  
XX PF 30-AUG-2000; 2000WO-US23735.  
XX  
XX PR 31-AUG-1999; 99US-0151575.  
XX  
XX PA (DUPO ) DU PONT DE NEMOURS & CO E. I.  
XX  
XX PA (PION-) PIONEER HI-BRED INT INC.  
XX  
XX PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;  
XX  
XX PI Simmons CR, Weng Z;  
XX  
XX DR WPI, 2001-244407/25.  
XX  
XX DR N-PSDB; AAS01165.  
XX  
XX PT New plant fertilization independent endosperm protein for the  
XX  
XX production of seed without fertilization is recombinantly produced -  
XX  
XX PS Claim 1; Page 54-55; 94pp; English.  
XX  
XX CC The sequence represents the amino acid sequence of fertilisation-  
XX  
XX independent endosperm protein, encoded by cDNA clone p0003.cgpf34rb.  
XX  
XX CC Fertilisation-independent endosperm proteins are plant reproduction  
XX  
XX proteins necessary for apomixis, the formation of seeds without  
XX  
XX fertilisation. Apomixis is especially useful to agriculture because it  
XX  
XX eliminates the necessity of selfing plants to produce genetically  
XX  
XX identical seed. Such seed is useful to produce seeds for human and animal  
XX  
XX food and for commercial milling and extraction, including the production  
XX  
XX of useful recombinant products in the endosperm e.g. pharmaceutical,  
XX  
XX nutraceutical, industrial compounds and polymers. Embryoleless seed  
XX  
XX production by transgenic plants is less likely to cause ethical and  
XX  
XX environmental concern over transgenic plant production as no gametes are  
XX  
XX being produced to cross pollinate with other crops and the seeds cannot  
XX  
XX germinate.  
XX  
XX SQ Sequence 273 AA;

Query March	55.98	Score 1103	DB 22	Length 273
Best Local Similarity	74.33	Pred. No. 7e-114		
Matches 197	Conservative 30	Mismatches 38	Indels 0	Gaps
OY	103	AGCYKGIIRVIDVNSERTHKSLVGHGDSVNEIRTPOLKPOLVITASKDESVRLMNVETGI	162	
Db	2	asngngllirvncatetkrlaksvfyghgdsineirtpklkpslllsaskdesvrlmnhvtll	61	
OY	163	CILFFAGAGGRYRVLSVDFPHSDIYFPFASGCMPTTIKINSMEKFFMYVEKSFPTWDPDS	222	
Db	62	cillffagaggrhnevlsvdfphsdierrffasgmdntrtkismkewfilyvdkssywtcllps	121	
OY	223	KFFPKFVQFPVFTASIRHTNYVDCNRMFGEDTILKLSVNDNEILWEPOLKENSPEGASDVL	282	
Db	122	kfpkfyvqfpvlraahvsnvydcrrwtlgtldllsksvnhelvwepkckegspgagsldll	181	
OY	283	LRYVPMCDIMFIRFSCDLHLSAASVAGOEKRYVWMDLSCPPVLIRKLSHNOSKSVIRO	342	
Db	182	qkypvpedvifrkfscdfhngqlaignrqklylvewegsapvllarlryngqskspirg	241	
OY	343	TAMSVYDSTILLACCEBDGTIMRWVYI	367	
Db	242	tavsfqstlllqagedgtlrmwdev	266	

RESULT	15
AAU00320	
ID	AAU00320 standard; Protein; 430 AA.

DT 12-SEP-2001 (first entry)

DE Fertilisation-independent endosperm protein, p0041.crtaw93r.

KW Fertilisation-independent endosperm; plant reproduction; apomixis.  
KW seed; pharmaceutical; nutraceutical; polymer; p0041.crtaw93r.

OS Zea mays.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

PN W0200116325-A2.

PD 08-MAR-2001.

PF 30-AUG-2000; 2000WO-US23735.

PR 31-AUG-1999; 99US-0151575.

PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
PA (PION-) PIONEER HI-BRED INT INC.

PI Butler KH, Danilevskaya O, Miao G, Morgante M, Sakai H;

AA  
DB  
WPT: 2001-244407/25

DR N-PSDB; AAS01167.

PT New plant fertilization independent endosperm protein for the production of seed without fertilization is recombinantly produced -

PS Claim 1; Page 56-57; 94pp; English.

CC The sequence represents the amino acid sequence of fertilisation-  
CC independent endosperm protein, encoded by cDNA clone p0041.critw9r.  
CC Fertilisation-independent endosperm proteins are plant reproduction  
CC proteins necessary for apomixis, the formation of seeds without  
CC fertilisation. Apomixis is especially useful to agriculture because it  
CC eliminates the necessity of selfing plants to produce genetically  
CC identical seed. Such seed is useful to produce seeds for human and animal

CC food and for commercial milling and extraction including the production  
CC of useful recombinant products in the endospore e.g. pharmaceutical,  
CC nutraceutical, industrial compounds and polymers. Embryonless seed  
CC production by transgenic plants is less likely to cause ethical and  
CC environmental concern over transgenic plant production as no gametes are  
CC being produced to cross pollinate with other crops and the seeds cannot  
CC germinate.

Sequence 430 AA;

Query Match	53.0%	Score 1045;	DB 22;	Length 430;
Best Local Similarity	51.8%	Pred. No. 4e-107;		
Matches 188;	Conservative 59;	Mismatches 84;	Indels 32;	Gaps 2

QY 6 LGNESIVGSLTPSNKSKYKVTNRIQEGKKPL YAVVFENELDARFEDVETAGNRITLYNC 65

Db 68 1gcevveg1lvpsrkreykpskysvtvgnhplyalgfnfidmrydvfaiaascnsvilyrc 127

QY 66 LGDAISALQSYADEDEKSEFYTVSWACG-VNGNPYVAAAGVKGIIRVIDVNSETIHKSL 124

Db 128 lenggfgllqnyvdedkdesfytllswtidqvdspllaagsnrilrvincateklksl 187

QY 125 VGHGDSVNEIRTOPLKPQLVITASKDESVRLWNVETGICILIFAGAGGHRREVLSDVDFHP 184

Db 188 vghgsihelrthaskpsliisask-----dfhp 216

QY 185 SDIYRFASCGMDTTIKIWSMKEFWTYVEKSFWTWDDPSKEPTKFVQFPVETASIHNTNYVD 244

Db 217 tevg1fascgmdntvkiwsmkefwilyveksyswtgphskfptrniqfpv1taavhsdyvd 276

QY 245 CNRWFGDFILSKSVDNELLWEPQLKENSPEGASDVLLRYPVPMCDIWFIFKSCDLHLS 304

Db 277 ctrwlgdflsksvknavllwepkpkrrpgegsvdvlqkypvpkchlwfmkfscdfysn 336

QY 305 SVAIGNQEGKVYVMDLKSCPEVLITKLSHNQSKSVIRQTAMSDVGSTILACCEDGTIWRW 364

Db 337 qmaignkgeiyvewqsspv1ldrlcngeckspirqtavsfdstilgaaddgaiwrw 396

QY 365 DVI 367

Db 397 dev 399

Search completed: June 25, 2002, 01:13:53  
Job time: 3737 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 25, 2002, 00:17:26 ; Search time 73.23 Seconds

(without alignments)  
484.186 Million cell updates/sec

Title: US-09-812-283-4

Perfect score: 1973

Sequence: 1 MSKITLGNESTVGSLLTPSNK.....STLLACCEDGTIRMDVITK 369

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR.71:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	715	36.2	425	2	S58672
2	217.5	11.0	1189	2	A12493
3	212.5	10.8	332	2	T29827
4	211.5	10.7	701	2	T16607
5	211.5	10.7	1227	2	AE1810
6	210.5	10.7	1356	2	T18521
7	202.5	10.3	704	2	S33263
8	202	10.2	559	2	AB2202
9	197.5	10.0	920	2	C96831
10	195	9.9	1526	2	AC2239
11	194.5	9.9	926	2	G96563
12	194.5	9.9	1747	2	AC1842
13	193	9.8	1258	2	A12155
14	188	9.5	798	2	S34023
15	187.5	9.5	421	2	T14773
16	183.5	9.3	586	2	T38992
17	183	9.3	1711	2	AD1842
18	182	9.2	606	2	T08180
19	181.5	9.2	380	2	T02085
20	181	9.2	1683	2	AF2071
21	179.5	9.1	518	2	B48088
22	176	8.9	377	2	T07376
23	176	8.9	616	2	C86239
24	175.5	8.9	1146	2	A55532
25	174.5	8.8	876	2	T31507
26	172.5	8.7	380	2	T03765
27	172	8.7	934	2	AG1889
28	170.5	8.6	515	2	S19487
29	170	8.6	377	2	T04089

30	169.5	8.6	304	2	AG1837	WD-40 repeat prote
31	169	8.6	703	2	T43557	F-box/WD-repeat pr
32	168.5	8.5	376	2	T19266	hypothetical prote
33	168	8.5	377	2	T04086	GTP binding protei
34	168	8.5	409	2	S36113	LIS-1 protein - hu
35	168	8.5	410	2	S48052	platelet-activatin
36	166.5	8.4	1551	2	AB2410	WD-repeat protein
37	166	8.4	375	2	T03256	GTP-binding protei
38	166	8.4	502	2	T41148	trp-asp repeat con
39	165	8.4	310	2	T43158	probable GTP-bind
40	164.5	8.3	326	2	T16987	GTP-binding protei
41	164.5	8.3	494	2	T19550	hypothetical prote
42	163.5	8.3	642	2	T39490	transcription init
43	163	8.3	713	2	JN0133	WD-40 repeat regul
44	162.5	8.2	240	2	T02059	GTP-binding regula
45	162.5	8.2	906	2	S35342	Golgi-associated p

## ALIGNMENTS

RESULT 1  
S58672  
extra sex combs protein - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 15-Feb-1996 #sequence-revision 01-Mar-1996 #text-change 26-May-2000  
C:Accession: S58672  
R:Gutjahr, T.; Freil, E.; Spicer, C.; Baumgartner, S.; White, R.A.H.; Noll, M.  
A:Title: The polycomb-group gene, extra sex combs, encodes a nuclear member of the WD  
EMBO J. 14, 4296-4306, 1995  
A:Reference number: S58672; MUID:96016202  
A:Accession: S58672  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-425 <GUT>  
A:Cross-references: GB:141867; NID:g1050936; PIDN:AAA86427.1; PID:g1050937  
C:Genetics:  
A:Gene: FlyBase:esc  
A:Cross-references: FlyBase:FBgn0000588  
C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match	Best Local Similarity	36.2%	Score 715;	DB 2;	Length 425;
Matches 143;	Conservative 76;	Mismatches 129;	Indels 16;	Gaps 10;	
QY 17 PSNKKSKYKYNRIQEGK-KPLVAVNFELDARFF-DVFTAGGNRTILYNCIGDAISAL 74					
DB 56 PKSRAYKXKTDHVKENHGANIFGVAFNTLLGKDEPQVFAFGSNRYTYECPRGQGMQLL 115					
QY 75 QSTADEKRESEFYTVSMAGV-NGNPYVAGGVKGIIRIVIDVNSEFTIHKSLVGHGDSVNE 133					
DB 116 HCYADDPDEVEFTYCAMSYDLKTSSPLAAGRGVIRIVDQNRVANVNYIGHGQAIN 175					
QY 134 IRRQPLKPOLVITASKDESVRLNVEYICILIFACAGGRVYVLSVDHPSPDYFASG 193					
DB 176 LKHPHKLILLSSGSKDHAIRLNNIOSHVCITALLGVGEGRDVLSIDFRMRG-DRIVSS 234					
QY 194 GMPYTKIKISMK--EFWTVYEKSFYTWDDPS--KPTKFEQFPVF--TASIHNTYVVCNRM 248					
DB 235 GMDHSIAKLMLCNTPPEFHKKIELSNITSQEKSTLPFTYTVKHPDFESTRIHRYVYVCVOM 294					
QY 249 FGDFTLSKSVNDEILMER-QIKEN---SPGASVDLLRYPVPMCDIWFIFKSCDLHL 303					
DB 295 FGFVYLSKSCENALVCMKPGQLHOSFEQVPPSDSCSTIAEFYDCEIWFVFGFENPMQ 354					
QY 304 SSVATINOGSKYVYVMDLCKSP--VLITKLSHNSQSVYROTMSVDSGSTIACCGDGTI 361					
DB 355 KYVALNQGQKYVWELDPSPDEGAMHTLLHNSRSVATVROIAFSRDSVATVVCODATV 414					
QY 362 WRWD 365					
DB 415 WRWN 418					

## RESULT 2

A12493

WD-repeat protein [imported] - Anabaena sp. (strain PCC 7120) plasmid pCC7120alpha  
 C:Species: Anabaena sp.  
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
 C:Accession: A12493

R:Kaneko, T.; Shimizu, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: A12493

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1189 &lt;KUR&gt;

A:Cross-references: GB:BA000020; PIDN:BA078213.1; PID:g17135667; GSPDB:GN00180

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr7129

A:Genome: plasmid

Query Match 11.0%; Score 217.5; DB 2; Length 1189;  
 Best Local Similarity 24.4%; Pred. No. 6.8e-10;

Matches 90; Conservative 57; Mismatches 149; Indels 73; Gaps 16;

QY 36 LIAVVEFLDARFDDVVTAG-GNRITLYNCLGDAISALQSYADEDEKESFYVSNACG 94

DB 691 VWSVVFHVG---QIATAGEDNTIKIMELQSGCCCLTQGHQHWKTIAFN----- 739

QY 95 VNGNPYVAAAGVKGIIIVIDVNSETIHKSLVGHGDSVNEIRTOPLPOLVITASKDESVR 154

DB 740 -SGGRILASGSFDONVLMIDHTGKCYMTLOGHTVTSVAFNP-KDNLISGSDOSVK 797

QY 155 LMNVEITGICILIFAGAGHREYLVSDPHPSDIYFASCGMDTTIKIWSM-----KEFW 208

DB 798 VMDKTRGCDLTJK--KHNRIMSVAFHPGQ-HLFVSGGDHAAKIMELGTGCTITFQ 853

QY 209 TYVEKST---WTDPSKPTKFFVQFPTASIHITNVYD-----CNRMF-- 249

DB 854 GHSNATYTIAMNMEHSLASGHEDQITKLMNLHSPKSNVNTHPRILOGHSNRFSV 913

QY 250 -----GFILISKSYDNEILMEP---QLKENSPEGASDVLRPVWMDIWFIFKSDCL 301

DB 914 VFSSTGQLLASGSADRTIKLMSPTGQCLHTLHGHSW-----VMAIAFSLDD 961

QY 302 HLSVAITAGNOEGKYYVNDLKS--CPVYLITKLSHNSQSVIROTAMSVDSGTTIACCEDG 359

DB 962 KL-LAGSYDHTVKIWDVSSGQC---LQTLQGHRS---VLAVAFSGDCKTLEFSSEYK 1013

QY 360 TIRWMDVIT 368

DB 1014 LVKQMDVET 1022

## RESULT 3

T29827

hypothetical protein C09G4.5 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T29827

R:Miller, N.; Gattung, S.  
 submitted to the EMBL data library, December 1995

A:Description: The sequence of C. elegans cosmid C09G4.

A:Reference number: Z20695

A:Accession: T29827

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-332 <MIL>

A:Cross-references: EMBL:U42438; PIDN:AAA83500.1; CESP:C09G4.5

C:Genetics:  
 A:Gene: CESP:C09G4.5  
 A:Introns: 65/3; 123/3; 185/2; 212/2

Query Match 10.8%; Score 212.5; DB 2; Length 332;

Best Local Similarity 22.4%; Pred. No. 3.2e-10;  
 Matches 79; Conservative 51; Mismatches 112; Indels 111; Gaps 15;

QY 32 GKRLVAVNEFLDARFDDVVTAGGNRITLYNCLGDAISALQSYADEDEKESFYVSN 91

DB 23 GKRF-----VLTAKLE-----DQKAIYGC-----AFQVAGDEQAVAVG- 62

QY 92 ACVGNPNYVAAAGVKGIIIVIDVNSETIHKSLVGHGDSVNEIRTOPLPOLVITASKDE 151

DB 63 -----GSFRLRSV-----GWEINDIRTCFANLNLIYCASDQ 94

QY 152 SVRLMNVETGICILIFAGAGHREYLVSDPHPSDIYFASCGMDTTIKIWSMKEFWT 209

DB 95 SIRIHIRNEACLVIGLGLCHAGETILSDVSTDDGE-ILSCGFHQLMEMDLVSKQVKE 153

QY 210 TYEKS-----FTWTD-----PSK-----FPTKE 228

DB 154 HLERACKALHODKINVLQSDIPYVSKGIMRSKSAVSRNPDKEDDLLEHRELIRPS 213

QY 229 VQFPVFTAS-----IHTNYYDCNRMF--GDFILSKSYDNE--ILLW-----EPQKENS 274

DB 214 CLIPYTPSSVSDMHSDDYDCIIRFLIGTNVYALSKGCGNEKAHFMRFPGPKGEVNRH 273

QY 275 G-----EGASDVLRIRVPMCDIWFIFKSCDLHLSVAIGNQESKYYVNDLKS 322

DB 274 GNVLRPKSCYTKTERTMNVPSGSAWFIKFAVDPRRRLVLCGAGSVAFEDLRN 326

## RESULT 4

hypothetical protein K10B2.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T16607

R:Miller, N.  
 submitted to the EMBL data library, June 1995

A:Description: The sequence of C. elegans cosmid K10B2.

A:Reference number: Z18545

A:Accession: T16607

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-701 <MIL>

A:Cross-references: EMBL:U28730; NID:9860694; PID:9860695; PIDN:AAA68258.1; CESP:K10B

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:K10B2.1

A:Introns: 78/3; 125/1; 183/2; 281/3; 404/3; 551/3; 668/3

Query Match 10.7%; Score 211.5; DB 2; Length 701;

Best Local Similarity 25.4%; Pred. No. 1.1e-09;  
 Matches 88; Conservative 55; Mismatches 118; Indels 85; Gaps 16;

QY 56 GGNRITLYNCLGDA--ISALQSYADE-----DKKE-----SFTYSMA 92

DB 244 GNYKMTIRINCQSENSKVVYCIQ-YDDKIVSGLDNNTIKIMDRDYSRILSGHTSVL 302

QY 93 CGVGNPNYVAAAGVKGIIIVIDVNSETIHKSLVGHGDSVNEIRTOPLPOLVITASKDES 152

DB 303 CLOYDNRYIISGSDAVYRWVDVETGECIKTLIHCAVHLR---FANGIMVTCSDRS 359

QY 153 VRLMNVETGICILIFAGAGHREYLVSDPHPSDIYFASCGMDTTIKIWSMKEFWTVE 212

DB 360 IAVMDVNSPRTITIRVLYGRAAVNVDF--DRIYVSAAG--DRIKIVSM----- 408

QY 213 KSFWTDDPSKPTKFFVQFPTASIHITNVYDCNRMFGDFILSKSYDNEILMEPQKEN 272



Db 409 -----DTLEF-----VRLAGHRGACLOQRGLVYVSGSSDNTIRLMDIH---- 449  
 Oy 273 SPGEASDVLLRY-----PVPMDIWFIFKSCDLHLSVAIGNQEKVYVMDLKS--PP 325  
 Db 450 -----SGCYLRYLEGHLELVC-----IRPD-----EKRIYSGAYDGKIKVMDLQALDPR 495  
 Oy 326 VLITKLSHNSKSVIRQTA-----MSVDSSTILACCEGDTIWRMDVI 367  
 Db 496 ALSSEIC-----LCSLVQHTGRVFRLOQDFDIYSSSHDITLIMDFL 538

RESULT 5  
 AE1810  
 WP-40 repeat protein [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.  
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
 C:Accession: AE1810  
 R:Kaneiko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriyuchi, Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AE1810  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1227 <KRP>  
 A:Cross-references: GB:BA000019; PIDN:BA077553.1; PID:g17135007; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr0029

Query Match 10.7%; Score 211.5; DB 2; Length 1227;  
 Best Local Similarity 25.1%; Pred. No. 2.3e-09;  
 Matches 82; Conservative 50; Mismatches 114; Indels 81; Gaps 15;

Oy 68 DGALISALQSVADDEKSESYTVS-----WACGVNGNP-----YVAGGVK 107  
 Db 697 DGRILA---SASDQITIKIMDITGNCQOTLIGHDWVMSVTFPTVTDPRLLASSSAD 753  
 Oy 108 GIIRIVDNSETIHKSLVGHGDSVNEIRTOPLKPOLVITASKDESVALNVENTGICILAF 167  
 Db 754 OHIKGLDVAATGCKLTGKHTREHVSFSP--DQOTLASSGSDTVRLMDVKKGQCMQIF 812  
 Oy 168 AGAGGRVYLVSDPHSPDIYRASCAMDTTITKISMKF-----FWTYEKSFTYTDDP 221  
 Db 813 E---GHSKRVYSVRFSP--DQOTLASSGSDRSIKLMDIQGECVNTLMGHSSQYWAIAFSP 868  
 Oy 222 SKFPFTFVGFVPTASIHNNYVDCNRMFGDFILSKSVNDEILMEPOLKENSPEGASDV 281  
 Db 869 D-----GRTILSCSDDOTARLMD--VITGNS-----LNT 895  
 Oy 282 LRRYPVPMCDIMFIFKSCDLHLSVAIGNQEKVYVMDLKS--CPPVLITKLSHNSKSV 339  
 Db 896 LRGY---TRDYVASFSPDSQI---LASGRDYYTIGLMNKTKGCHP-----LGHQGR-- 943  
 Oy 340 IRTQAMSVDSSTILACCEGDTIWRMDV 366  
 Db 944 IRSVAFHPDGKILASSGSDNTIRKIMDI 970

RESULT 6  
 T18521  
 beta transducin-like protein - Podospora anserina  
 C:Species: Podospora anserina  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T18521  
 R:Saunpe, S.; Turcq, B.; Begueret, J.  
 Gene 162, 135-139, 1995  
 A:Title: A gene responsible for vegetative incompatibility in the fungus Podospora anserina  
 A:Reference number: Z18944; MUID:96009891  
 A:Accession: T18521

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1356 <SAU>  
 A:Cross-references: EMBL:L28125; NID:g607002; PID:g607003; PIDN:AAA85775.1  
 C:Genetics:  
 A:gene: het-el  
 A:introns: 761/3

Query Match 10.7%; Score 210.5; DB 2; Length 1356;  
 Best Local Similarity 28.2%; Pred. No. 3.2e-09;  
 Matches 78; Conservative 31; Mismatches 103; Indels 65; Gaps 11;

Oy 101 VAAGVKGIIIRIVDNSETIHKSLVGHGDSVNEIRTOPLKPOLVITASKDESVALNVENT 160  
 Db 940 VASGSDHTIKIMDAASGCTOTLEBGSSVLSVAFSP--DQGVASGSDGDKTIKIMDTAS 998  
 Oy 161 GICILIFAGAGHRYEVLSDPHSPDIYRASCAMDTTITKISMKFETWYVEKSFYTDD 220  
 Db 999 GCTGTGLEHGC---SVMSVAFSP--DQGVASGSDDKTIKIM-----DT 1038  
 Oy 221 PSKFPFK-----FYQFPYFTASIHNNYVDCNRMFGDFILSKSVNDEILMEP---OLK 270  
 Db 1039 ASGCTGTGLEHGCWQSVVFSFD-----GQRYASGSDHTIKIMDAVSGTCY 1086  
 Oy 271 ENSPGEASDVLLRYPPVPMCDIMFIFKSCDLHLSVAIGNQEKVYVMDLKS--CPPVL 328  
 Db 1087 QLEHGSDS-----VMSVAFSP--GQRYASGSDHTIKIMDAASGCTOTL- 1131  
 Oy 329 TKLSHNSKSVIRQTAAMSVDSSTILACCEGDTIWRMD 365  
 Db 1132 -----EGHGGWVHSAFSPDQGVASGSDITGTIKIMD 1163

RESULT 7  
 S33263  
 transcription initiation factor IID-associated protein, 80K - fruit fly (Drosophila m  
 N:Alternate names: tightly-associated factor TAF-II-80  
 C:Species: Drosophila melanogaster  
 C:Date: 31-Mar-1992 #sequence\_revision 14-Sep-1994 #text\_change 01-Dec-2000  
 C:Accession: S33263; S45729; A54593  
 R:Dynlacht, B.D.; Weinzierl, R.O.J.; Admon, A.; Tjian, R.  
 Nature 363, 176-179, 1993  
 A:Title: The DRAF(II)80 subunit of Drosophila TFIID contains beta-transducin repeats.  
 A:Reference number: S33263; MUID:93247643  
 A:Accession: S33263  
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-704 <DN>  
 A:Accession: S45729  
 A:Molecule type: protein  
 A:Residues: 105-126;226-268;273-301;321-336 <DY2>  
 R:Fokkubo, T.; Gong, D.W.; Yamashita, S.; Takada, R.; Roeder, R.G.; Horikoshi, M.; Nak  
 Mol. Cell. Biol. 13, 7859-7863, 1993  
 A:Title: Molecular cloning, expression, and characterization of the Drosophila 85-kil  
 A:Reference number: A54593; MUID:94067146  
 A:Contents: Schneider cells, embryos  
 A:Accession: A54593  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid; protein  
 A:Residues: 1-409; 'KL' 412-704 <KOK>  
 A:Note: sequence extracted from NCBI backbone (NCBIP.139815)  
 C:Genetics:  
 A:gene: FlyBase:Taf80  
 A:Cross-references: FlyBase:FBgn0010356  
 A:Map position: 47C5-6  
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology  
 C:Keywords: transcription initiation  
 F:374-406/Domain: WD repeat homology <WD1>  
 F:446-479/Domain: WD repeat homology <WD2>  
 F:488-521/Domain: WD repeat homology <WD3>  
 F:530-563/Domain: WD repeat homology <WD4>  
 F:572-605/Domain: WD repeat homology <WD5>

F;614-647/Domain: WD repeat homology <WD6>

**Query Match**      **Similarity**    10.3%; Score 202.5; DB 2; Length 704;  
Best Local Similarity    25.28; Pred. No. 6.2e-09;  
**Matches**    65; Conservative    49; Mismatches    99; Indels    45; Gaps    13;

Dy    110 IRVIDNSEFIHKSIVGHGDSVNEIRTPOLKPOLVTITASKDESVRLMNVETICILIFAG 169  
         :|::|| :::||:|||| |:: ||:: ||: ||: ||: ||: ||:  
Db    432 VRHLDRSGEVRTSLMHGTGPYRCACAF--EMNLLSCSEGSTIKLMLLTWMSCVTYR- 489  
Dy    170 AGHRREVLVSDFHPSPDIYRFASCGMTTIIKWSKKEWTVEKSFYTWDPSKFPTKY 229  
         ||:: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db    490 --GHVPVMVDVRAPHGXY-FVSCSYDKTLARLA-----TDNQ----- 525  
Oy    230 QPFFVFASHITNTVDCNRWF--GDFTLSKVNDIELLEMPOLKENSPGCASDVILLRPV 287  
Db    526 ALRVFVG--HLSDVCQCFPHNSNYVAATGSDDRYLRM----DNMGTSVSR-LMIIGHG 577  
Oy    288 PMCDIPIFKPCDLHLSSVAIGNOEGRVYMDLKSCPVLITKLISHNOFSVIROTAMSV 347  
Db    578 SVSSLAE--SACGRYLAA---GSVDHNLIIMWL-SNGSLVTLLRHRT---STVTITFR 628  
Oy    348 DGSTILIACCEDGITWRMD 365  
         ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db    629 DGVTLAAGAIGLDNNLTLMWD 646

**RESULT**         8  
AB2202  
hypothetical protein allj3169 [imported] - Anabaena sp. (strain PCC 7120)  
C.Species: Anabaena sp.  
A.Note: AAnabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C.Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
CAccession: AB2202  
R.Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matenabe, A.; Iriyuchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A.Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A.Reference number: AB1807; MUID:21595285; PMID:11759840  
A.Accession: AB2202  
A.Status: Preliminary  
A.Molecule type: DNA  
A.Residues: 1-559 <KUR>  
A.Cross-references: GB:BA000019; PIDN:BAB74868.1; PID:g17132264; GSPDB:GN00179  
A.Experimental source: strair FCC 7120  
C.Genetics:  
A.Gene: allj3169

**Query Match**      **Similarity**    10.2%; Score 202; DB 2; Length 559;  
Best Local Similarity    25.3%; Pred. No. 5e-09;  
**Matches**    73; Conservative    50; Mismatches    116; Indels    50; Gaps    12;

Oy    86 FYYTVMACGVANGNPYAAGVKGIIRVIDVNSETIHKSLVGHGDSVNEIRTOPLKPOLI 145  
         :|:: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db    270 YSTNSLAISPDGKN-IASGGDGKIIRLEMINTOKFLACFGSHSOAVTSVFP-OGCIL 327  
Oy    146 TASNDESURLMNNETVICILIFAGAGCHREVELVSDFHSIDIRFASCGMDTTIKISM 205  
         ||:: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db    328 TASDDTKTIKMHLPPISSEFTL---NGHTNPVKSVSFPNG-OILIASGWKKOVKMWLV- 382  
Oy    206 EFWTVYEKSWTWDPDSKPTKFEVOPPVETAJIHTNYVCNMRFGEFILSKSVNDELTM 265  
         |: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db    383 -----TGKETLYALKAHQLQSAVAFAFPQ-----GETLASASPDRITRLWM 422  
Oy    266 EPOLKENSPGCASDVILLRPVPMK-----DWTFKSCDLHLSVAIGNOEGRVYMD 319  
         |: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db    423 --QTQNHF-----RYTLIKTLSGHTRAVLAIAFSPDGRI--LATGSDDTIKLMD 469  
Oy    320 LKSCCPVLITKLISHNOFSVIROTAHSVDSGITLACCEEGTITRMADYIT 368  
         :|:: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
Db    470 INT-GOIATLTLGHSMWSVVAALFTA--DNKTILSIASMKTIKLMKYST 514

RESULT 9  
 C96831  
 hypothetical protein F18B13.7 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
 C:Accession: C96831  
 R:Ethnologists: A.; Ecker, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.; Alon  
 Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
 ansen, N.F.; Hughes, B.; Hutzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malti, R.; Merzla  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.W.; Sun, H.; Tallio  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: AB6141; MUID:21016719  
 A:Accession: C96831  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-920 <STD>  
 A:Cross-references: GB:AE005173; NID:g5902363; PIDN:AD5465.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F18B13.7  
 A:Map position: 1  
 C:Superfamily: coatomer complex beta' chain; WD repeat homology

RESULT 10  
AC2239  
WD-40 repeat protein [imported] - *Anabaena* sp. (strain PCC 7120)  
C.Species: *Anabaena* sp.  
A.Note: *Anabaena* sp. (strain PCC 7120) is a synonym of *Nostoc* sp. strain PCC 7120  
C.Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C.Accession: AC2239  
R.Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriki, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takahawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A.Title: Complete genomic sequence of the filamentous nitrogen-fixing *Cyanobacterium*  
A.Reference number: AB1807; MUID:21595285; PMID:11759840  
A.Accession: AC2239  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-1526 <KUN>  
A.Cross-references: GB:BA000019; PIDN:BA875165.1; PID:G17132599; GSPDB:GN00179  
A.Experimental source: strain PCC 7120  
C.Genetics:  
A:Gene: alr3466

Query	Match	9.9%	Score 195;	DB 2:	Length 1526;
	Match Similarity	26.6%	Pred. No. 7.6e-08;		
	Matches	79;	Conservative 37;	Mismatches 107;	Indels 74;
					Gaps 14;
OY	90	SMACVGNCP---	YAAAGVKGIIIRVIDVNSEFTIHKSLVGHGDSVNELRTOQLKRLQVLT	146	
Db	1243	SWNSVVENPDGSMLASGSSDKTYVLMWIDISSKCLHTFOGHTHNNVNSVAFN-DGSMLAS	1301		
OY	147	ASKESEVILAMNVEIGICILIFAGAGHREYLVSDFHESDIYRFASCGMDTITIKIWSMK	206		
Db	1302	GSQDOFVRLMEISSLKCLHTFO---GHTSWWSVTFSP-DGMLASGSDDOQVRLMSTISS	1357		
OY	207	--FWTYVEKSPFTWDDSPKFPFKFQFPVFATSIHTNYVDCNRMGDFILSK-----	256		
Db	1358	GECLTYTFIG-----HTN-----WVGSVTFSPDGALLAS	1385		
OY	257	-SYDNEILLMEPQLKENSPGEGASDVLLRYPYPCMDIWF--IKFSCDLHLSSVAIGNOG	313		
Db	1386	GSQDOFVRLM-----SISGSKCLTYLQGHNNWGVISVFPDGTL--LASGSDQ	1432		
OY	314	KYYVMDLKS--CPVLTIRKLSHNSOKSVIRQTAMVSGDSTILACCEDTIRMMWDYIT	368		
Db	1433	TYRLMNISSGEC---LYTLHGHTNS---VRSVAEFSDDGLILASGSDDETITKIMWDYT	1483		

RESULT 11  
G96563  
probable coatomer complex subunit, 33791-27676 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: G96563  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, T.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Liu, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marzialis,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141, MUID:21016719  
A:Accession: G96563  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1..926 <STD>  
A:Cross-references: GB:AE005173; NID:G10645445; PIDM:AMG21561.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F19K6.16  
A:Map position: 1  
A:Superfamily: coatomer complex beta' chain; WD repeat homology

	Query Match	9.9%	Score 194.5:	DB 2:	Length 926:
	Best Local Similarity	26.2%:	Pred. No. 4.2e-08:		
	Matches	60;	Conservative	35;	Mismatches 81; Indels 53; Gaps 9.
OY	100 YVAGGVKGIIRVIDYNSETIHKSLVGHDSDVNEIRTOPLKPOLYITASKDESVRLMNYE	159			
	:        :    :				
Db	71 WVAAGADMIVRYNXYNTMDKVKEFAHSDYIRCAVAHPTEP-YVLSSDDMLIKLDMWE	129			
	:        :    :				
OY	160 TG-IQILIFAGGGRHYELVDPIRPSDIYRASCGMTPTIKINMSKEFWTVYSFTYT	218			
	:        :    :				
Db	130 KGMACTQLPE--GSHSYVMQVTENPKDITNFASASLDPTIKIWNL-----GS	174			
	:        :    :				
OY	219 DDPSEFPTREVOVPVFETIASHTNYDCNMWF--GG--FLTSKYNDNEILLMEPOLKE---	271			
	:        :    :				
Db	175 PDPN-----FTLDAHQGKNVCVDYFFLGGRKPYLTISDSDDHTAKVMVQYQRKSQVQ	223			
	:        :    :				
OY	272 --NSFGEGASDYLLRPVPMCDIMWTFKSCDLHLSSVALIGNOEGKKYYW	318			
	:        :    :				
Db	224 TLEGTHNNVASVCFHPDELPI-----IIIGSDDGYVRWI	256			
	:        :    :				

RESULT 12  
 AC1842  
 WD-40 repeat protein [imported] - *Anabaena* sp. (strain PCC 7120)  
 C:Species: *Anabaena* sp.  
 A:Note: *Anabaena* sp. (strain PCC 7120) is a synonym of *Nostoc* sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
 C:Accession: AC1842  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu  
 Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AC1842  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1747 <KRR>  
 A:Cross-references: GN:BA000019; PIDN:BA877807.1; PID:G17135261; GSPDB:GN00179  
 C:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: al10283

[illegible]

RESULT 13

AT12155

WD-repeat protein [Imported] - *Anabaena* sp. (strain PCC 7120)

C:Species: *Anabaena* sp.

A:Note: *Anabaena* sp. (strain PCC 7120) is a synonym of *Nostoc* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002

C:Accession: AT12155

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Itriguen, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata DNA Res. 8, 205-213, 2001

A:Title: Complete Genome Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AT12155

A:status: preliminary

A:molecule type: DNA

A:Residues: 1-1258 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA674499.1; PID:G17131893; GSPDB:GN00179







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 25, 2002, 01:15:16 ; Search time 37.87 Seconds  
(without alignments)  
377.278 Million cell updates/sec

Title: US-09-812-283-4

Perfect score: 1973  
Sequence: 1 MSKITIGNESTVGSITPNSK.....STILACCEDGTIMRWIVTK 369

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : SWISSPROT\_40:\*  
Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	211.5	10.7	665	1 LI23_CAEEL	Q09990 caenorhabd1
2	210.5	10.7	1356	1 HET1_PODAN	Q00808 podopora a
3	202.5	10.3	704	1 T2D4_DROME	P49846 drosophila
4	194.5	9.9	542	1 FW1B_HUMAN	Q9ukb1 homo sapien
5	193	9.8	1258	1 Y500_ANASP	O8ylc2 anabaena sp
6	188	9.5	798	1 T2D4_YEAST	P38129 saccharomyc
7	188	9.5	914	1 COPP_DROME	O62621 drosophila
8	183.5	9.3	586	1 TUI2_SCHPO	O9unq8 schizosacch
9	182	9.2	606	1 PF20_CHLRE	P93107 chlamydomon
10	181.5	9.2	380	1 GBB_MAIZE	P49178 zea mays (m
11	179.5	9.1	518	1 TRCB_XENLA	O91854 xenopus lae
12	177.5	9.0	605	1 FW1A_HUMAN	O9y297 homo sapien
13	176	8.9	377	1 GBB_SOLTU	P49695 solanum tub
14	176	8.9	742	1 PKWA_THECU	P49695 thermomons
15	175.5	8.9	1146	1 KMAA_DICDI	P42527 dictyosteli
16	175	8.9	316	1 GBLP_BIOGL	O93134 biopalmari
17	173	8.8	1249	1 APAF_RAT	O9epv5 rattus norv
18	172.5	8.7	380	1 GBB_ORYSA	O40687 oryza sativ
19	171	8.7	318	1 GBLP_DROME	O18640 drosophila
20	170.5	8.6	515	1 YCW2_YEAST	P25852 saccharomyc
21	170	8.6	377	1 GBB2_TOBAC	P93398 nicotiana t
22	169.5	8.6	318	1 GBLP_TRYBB	O94775 trypanosoma
23	169	8.6	703	1 POP2_SCHPO	O14170 schizosacch
24	168.5	8.5	904	1 YK4_CAEEL	Q17963 caenorhabd1
25	168.5	8.5	376	1 COPP_RAT	O35142 rattus norv
26	168	8.5	377	1 GBL1_TOBAC	P93397 nicotiana t
27	168	8.5	409	1 LIS1_BOVIN	P43033 bos taurus
28	168	8.5	409	1 LIS1_HUMAN	P43034 homo sapien
29	168	8.5	409	1 LIS1_MOUSE	P43035 mus musculu
30	167.5	8.5	800	1 T2D4_HUMAN	O15542 homo sapien
31	166	8.4	375	1 GBB3_TOBAC	Q40505 nicotiana t
32	166	8.4	1249	1 APAF_MOUSE	O88879 mus musculu
33	165.5	8.4	514	1 TUP1_CANAL	P56093 candida alb

34	164.5	8.3	317	1 GBLP_HYDAT	Q25189 hydra atten
35	164.5	8.3	326	1 GBLP_NICPL	P93340 nicotiana p
36	163	8.3	332	1 GBLP_DICDI	P46800 dictyosteli
37	163	8.3	713	1 TUP1_YEAST	P16649 saccharomyc
38	162.5	8.2	905	1 COPP_BOVIN	P35605 bos taurus
39	162.5	8.2	905	1 COPP_HUMAN	P35606 homo sapien
40	162.5	8.2	905	1 COPP_MOUSE	O55029 mus musculu
41	162	8.2	314	1 GBLP_SCHPO	O10281 schizosacch
42	160.5	8.1	317	1 GBLP_ORENI	O42249 oreochromis
43	160.5	8.1	377	1 GBB_ARATH	P49177 arabidopsis
44	159.5	8.1	640	1 MT30_YEAST	P39014 saccharomyc
45	159	8.1	377	1 GBL_NICPL	P93339 nicotiana p

## ALIGNMENTS

RESULT 1  
LI23\_CAEEL STANDARD: PRT: 665 AA.  
AC Q09990: Q09990: 15-JUL-1998 (Rel. 36, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE F-box/WD-repeat protein lin-23.  
GN LIN-23 OR K10B2.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, DEVELOPMENTAL STAGE, AND MUTAGENESIS.  
RX MEDLINE=20515608; PubMed=11060233;  
RA Kipreos E.T., Gohei S.P., Hedgecock E.M.;  
RT "The Caenorhabditis elegans F-box/WD-repeat protein lin-23 functions  
RT to limit cell division during development.";  
RL development 127:5071-5082(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Miller N.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Functions cell autonomously to negatively regulate cell  
CC cycle progression. Required to restrain cell proliferation in  
CC response to developmental cues. Probably recognizes and binds to  
CC some proteins and promotes their ubiquitination and degradation  
CC (By similarity).  
CC -!- SUBUNIT: Part of a SCF (SKP1-cullin-F-box) protein ligase complex  
CC (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -!- DEVELOPMENTAL STAGE: Highest levels in embryos and adults, lowest  
CC levels in larvae. Maternal expression results in high zygotic  
CC levels.  
CC -!- SIMILARITY: CONTAINS 1 F-BOX DOMAIN.  
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TTP-ASP DOMAINS).  
CC -!- SIMILARITY: STRONG, TO X LAEVS FAXWI.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF275253; AAG28037.1; -  
CC EMBL: U28730; AAA68258.2; -  
CC WormPep: K10B2.1; CE28600.  
CC InterPro: IPR001810; F-box.





DT 01-OCT-1996 (Rel. 34, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Transcription initiation factor TFIID 85 kDa subunit (P85) (TAFTII-80).  
 GN TA80 OR CG7704.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 NX NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 105-120, 272-280 AND 365-374.  
 RC TISSUE=Embryo;  
 RX MEDLINE=94067146; PubMed=8247000;  
 RA Kokubo T., Gong D.-W., Yamashita S., Takada R., Roeder R.G.,  
 RA Horikoshi M., Nakatani Y.;  
 RT "Molecular cloning, expression, and characterization of the  
 RT Drosophila 85-kilodalton TFIID subunit.";  
 RL Mol. Cell. Biol. 13:7859-7863(1993).  
 [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93247643; PubMed=8483503;  
 RA Dynlacht B.D., Weinzierl R.O.J., Admon A., Tjian R.;  
 RT "The TFIID subunit of Drosophila TFIID contains beta-transducin  
 RT repeats.";  
 RL Nature 363:176-179(1993).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,  
 RA Beeson K.Y., Bensu P.V., Bernier B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein F., Brotlier P.,  
 RA Butris K.C., Busam D.A., Butler H., Cadelau E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
 RA Flier C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jajani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paolel J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 [4]  
 RP FUNCTION: TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A  
 RP CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS  
 RP AND REPRESSORS. TFIID-80 INTERACTS DIRECTLY WITH TBP AND TAFII-110.  
 RP MAY PLAY A ROLE IN HELPING TO ANCHOR TAFII-110 WITHIN THE TFIID  
 RP COMPLEX. MAY BE INVOLVED IN TRANSDUCING SIGNALS FROM VARIOUS

CC TRANSCRIPTIONAL REGULATORS TO THE RNA POLYMERASE II TRANSCRIPTION  
 CC MACHINERY.  
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A  
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).  
 CC -1- SIMILARITY: TO YEAST TAFII-90 AND MAMMALIAN TAF2D.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: U06460; AAC6481.1; -.  
 CC EMBL: AE003828; AAF58737.1; -.  
 CC TRANSFAC: T02122; -.  
 CC DR FLYBASE: FBgn010356; TAF80.  
 CC DR InterPro: IPR001680; WD40.  
 CC DR Pfam: PF00400; WD40; 6.  
 CC DR PRINTS: PR00320; GPROTEINBRPT.  
 CC DR SMART: SM00320; WD40; 6.  
 CC DR PROSITE: PS00678; WD\_REPEATS\_1; 3.  
 CC DR PROSITE: PS50082; WD\_REPEATS\_2; 6.  
 CC DR PROSITE: PS50294; WD\_REPEATS\_REGION; 2.  
 CC DR Transcription regulation: Nuclear protein; Repeat; WD repeat.  
 CC KW REPEAT  
 CC FT REPEAT 376 405 WD 1.  
 CC FT REPEAT 448 478 WD 2.  
 CC FT REPEAT 490 520 WD 3.  
 CC FT REPEAT 532 562 WD 4.  
 CC FT REPEAT 574 604 WD 5.  
 CC FT REPEAT 616 646 WD 6.  
 CC FT CONFLICT 410 411 KL -> NV (IN REF. 2).  
 CC SQ SEQUENCE 704 AA; 79324 MW; F4E2F12534501EF6 CRC64;  
 Query Match 10.3%; Score 202.5; DB 1; Length 704;  
 Best Local Similarity 25.2%; Pred. No. 2.2e-09;  
 Matches 65; Conservative 49; Mismatches 99; Indels 45; Gaps 13;  
 QY 110 IVIDIVNSITKSHVGHDSVNEIRTOPLKPOLVITASKDSVRLMANETGICILIFAG 169  
 DB 432 VMLDNRSGEVTRSLMGHTGPRYRCAPAP-EMNLISCSDSTIRKMSLLTSCVYTYR- 489  
 QY 170 AGCHREYVLVDPHSPSDIYRFASCGMDTPIKITSKMEFTYVEKSFVTWDDPSKEPTKEV 229  
 DB 490 --GHYVPWVDVFAFPHGYV-FVSCSYDKTARLMA-----TDSNQ----- 525  
 QY 230 QRPVFLASHTHTVYDCNRKF--GDPLFSVYNEILLMPOLKENSFGCASDYLIRPV 287  
 DB 526 ALRVFVG--HLSDVDCVQPHNSNVATGSSDRTVRLW-----DNMTGQSVR-LMTGHNG 577  
 QY 288 PMCDLWIFKSCDLHLSVAIGNOEGKYVWDLPKSCPPVLITKLSHNSKQSVYRTAMSV 347  
 DB 578 SVSSLAFL--SAGGRYLA---GSVDHNIITMDL-SNGSLVTLIRHT---SVYITITFSR 628  
 QY 348 DSGTILACCEDEGTIWRMD 365  
 DB 629 DGTVLAAGLDNNLTFLMD 646  
 RESULT 4  
 FW1B\_HUMAN STANDARD; PRT; 542 AA.  
 AC Q9UBK1: Q9Y4C6: Q9P2S8: Q9P2S9:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, last sequence update)  
 DT 01-MAR-2002 (Rel. 41, last annotation update)  
 DE F-box/WD-repeat protein 1B (F-box and WD-repeats protein beta-Trip2).  
 GN FBXW1B OR FBW1B OR BTCP2 OR KIAA0696.  
 OS Homo sapiens (Human).

```

CC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20003060; PubMed=10531035;
RA Cenciarelli C., Chlaud D.S., Guardavaccaro D., Parks W., Vidal M.,
RA Pagano M.;
RT "Identification of a family of human F-box proteins.";
RL Curr. Biol. 9:1177-1179(1999).
RN [2]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal lung;
RX MEDLINE=20160458; PubMed=10694485;
RA Koike J., Sagara N., Kirikoshi H., Takagi A., Miya T., Hirai M.,
RA Kaisho M.;
RT "Molecular cloning and genomic structure of the betaTRCP2 gene on
RL chromosome 5q35.1."
RL Biochem. Biophys. Res. Commun. 269:103-109(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RL The complete sequences of 100 new cDNA clones from brain which can
RL code for large proteins in vitro."
CC -1- FUNCTION: Probably recognizes and binds to some phosphorylated
CC proteins and promotes their ubiquitination and degradation.
CC -1- SUBUNIT: PART OF A SCF (SKP1-CULIN-F-BOX) PROTEIN LIGASE COMPLEX.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: A, B AND C (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 1 F-BOX DOMAIN.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, AF176022; AAF04528.1; -
DR EMBL, AB033279; BAA92329.1; -
DR EMBL, AB033280; BAA92330.1; -
DR EMBL, AB033281; BAA92331.1; -
DR EMBL, AB014596; BAA31671.1; ALT_INIT.
DR MIM, 605651; -
DR InterPro: IPR001810; F-box.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00646; F-box; 1.
DR Pfam: PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINRPT.
DR SMART; SM00256; FBOX; 1.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS50181; FBOX; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 5.
DR PROSITE; PS50082; WD_REPEATS_2; 7.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Ubiquitin conjugation; Repeat; WD repeat; Alternative splicing.
FT DOMAIN 129 167 F-box.
FT REPEAT 238 275 WD 1.
FT REPEAT 278 315 WD 2.
FT REPEAT 318 355 WD 3.
FT REPEAT 361 398 WD 3.
FT REPEAT 401 440 WD 4.
FT REPEAT 442 478 WD 5.
FT REPEAT 490 527 WD 6.
FT REPEAT 527 564 WD 7.
FT VARSPLIC 16 49 MISSING (IN ISOFORM A).

```

FT	VARSPLIC	16	48	CSPVRSIMIGCANCTIVMCSLOSIMPVSRLC -> NRSV
FT	SEQUENCE	542 AA;	62090 MM;	MEQNDNEBSKNTM (IN ISOFORM B).
SO	SEQUENCE	542 AA;	62090 MM;	7CD40087EFAA5C8A CRC64;
	Query Match	9.9%;	Score 194.5;	DB 1; Length 542;
	Best Local Similarity	26.1%;	Prod. No. 7.3e09;	
	Matches 77;	Conservative	44;	Mismatches 107; Indels 67; Gaps 14;
OY	87	YTVSMACGVNCPNPPVAAAGVGGIIRVIDVNSETIHKSLVGHGDSVNEIRFOPLKPOLVIT	146	
DB	279	HTGSVLCLQDYDERIVLVGSSDSFVRWMDVNTGEVILNTLIHNEAVLHLR--FSNGLMVT	335	
OY	147	ASKDESVLWNVENGICILIFAGAGHRYEVLSDVDFHSDIYRRAASGMPTTIKWSMKE	206	
DB	336	CSKDRSLIYVMDASATDTTLRLRVLVGHRFAAVNVDF--DDKTIYASG-DRTIKVMS---	389	
OY	207	FMTYVKSFTWTDPSKPTKFOVPVETASIHNTYVDCNRFQDFILSKSDNEILLME	266	
DB	390	-----TSTCEE-VNTLNGHKRGIALQYRRIYVSGSSDNTIRLMD	429	
OY	267	POLKENSFG-----EGASDVLLRLYRPMCDIWFIKFSCDLHLSVLAIGNEGAVYVWL	320	
DB	430	IEC-----GACLRVLEGHELY-----RC-----IRFP-----NKRIVSGAYDGKIKWDL	470	
OY	321	KSC-----LPV-----LITKLHNSKSVIRQIRAMSVSDSTILACCEDGIRWMDVI	367	
DB	471	QAALDPRAPASTLCITLRTVEH--SERVER---LQDFEQITSSSHDPTITILWDFL	520	
	RESULT	5		
YS00	ANASP			
ID	YS00_ANASP	STANDARD:	PRT:	1258 AA.
AC	Q8YTC2;			
DT	01-MAR-2002 (Rel. 41, Created)			
DT	01-MAR-2002 (Rel. 41, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical WD-repeat protein alr2800.			
GN	ALR2800.			
OS	Anabaena sp. (strain PCC 7120).			
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.			
OX	NCBI_TaxID=103690;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=21595285; Pubmed=11759840;			
RA	Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,			
RA	Watanabe A., Iriuchih M., Ishikawa A., Kawashima K., Kimura T.,			
RA	Kisida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,			
RA	Nakazaki N., Shimoto S., Sugimoto M., Takazawa M., Yamada M.,			
RA	Yasuda M., Tabata S.;			
RT	Complete genomic sequence of the filamentous nitrogen-fixing			
RT	Cyanobacterium Anabaena sp. strain PCC 7120."			
RL	DNA Res. 8:205-213(2001).			
CC	- SIMILARITY: CONTAINS 15 WD REPEATS (TRP-ASP DOMAINS).			
CC	-			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: AP003590; BAB74499.1; -			
DR	PROSITE: PS00678; WD_REPEATS_1; 9.			
DR	PROSITE: PS50082; WD_REPEATS_2; 14.			
DR	PROSITE: PS50294; WD_REPEATS_REGION; 1.			
KW	Hypothetical protein; Repeat; WD repeat; Complete proteome.			
FT	REPEAT	55	93	WD 1.
FT	REPEAT	640	679	WD 2.
FT	REPEAT	682	721	WD 3.
FT	REPEAT	724	763	WD 4.
FT	REPEAT	766	807	WD 5.

```

FT REPEAT 809 849 WD 6.
FT REPEAT 850 889 WD 7.
FT REPEAT 892 931 WD 8.
FT REPEAT 934 975 WD 9.
FT REPEAT 976 1017 WD 10.
FT REPEAT 1019 1059 WD 11.
FT REPEAT 1060 1101 WD 12.
FT REPEAT 1103 1143 WD 13.
FT REPEAT 1144 1183 WD 14.
FT REPEAT 1186 1227 WD 15.
SQ SEQUENCE 1258 AA; 139513 MW; 45DF03B91170C451 CRC64;

Query Match 9.8%; Score 193; DB 1; Length 1258;
Best Local Similarity 23.2%; Pred. No. 3e-08;
Matches 87; Conservative 56; Mismatches 134; Indels 98; Gaps 17;

OY 55 AGNRRITL-----YNCIGDAISALOSYA-----DEKKESEFYVSMACVNGN 98
DB 588 AGGNVNLNLHNAQVDLSGDFSG---LTYWQAYLOGVNLHNDVFNDSLSCCVFTETLGN 644
OY 99 PYVAGGVK-----IIRVIDVNETHKSILVGHGDSVNEIRPOLKPOLVITAS 148
DB 645 ILTSAFSEGGOLATCDPDCDHVRWEVSKGLLLICRGHSNMYRFVYFSP-DGEILASCG 703
OY 149 KDESRRLNVTGICILIFAGAGHREYLSVDFHPSDIYRASCMDTTIKMSKFEW 208
DB 704 ADENVKLSVSRGVCIKITLT--GHEHEVSYAFHP-DGETLASGDKTILKMDIQD-G 758
OY 209 TYVEKSFMTDDPSKFEKFEVQFVFETASIHNYVDCNRMF--GDFILKSVDNETILWE 266
DB 759 TGLQ-----TLGHTDWRCAVAFSPDQNTLASADHTIKLMD 796
OY 267 PO-----LKNNSP-----GEGASDVLLR--YPPVPM-----DIW 293
DB 797 VSQGRCLRTKSHGTQWRSVAFSADQTLASGSDRTIKIMWYHGECKTYIGHNYSY 856
OY 294 FIKESCDLHLSVAIGNOGKRYVWDLKSCPVLTIKLSHNSKSVYRTTANSDGSTITL 353
DB 857 STAYSPDSKI--LVSGSGGRITKLMW--CQTHICIKTLHGHTNEVC-SVAFSPDQTLA 910
OY 354 ACCEDGTIRWMDVIT 368
DB 911 CYSIDQSVRLMNCRT 925

RESULT 6
T2D4_YEAST STANDARD: PRF: 798 AA.
AC P38129;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Transcription initiation factor TFIID 90 kDa subunit (TAFII-90).
GN TAF90 OR YBR198C OR YBR1410.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=95066384; PubMed=7975899;
RA Mallet L., Bussiereau F., Jacquet M.;
RT "Nucleotide sequence analysis of an 11.7 kb fragment of yeast
RT chromosome II including BEM1, a new gene of the WD-40 repeat family
RT and a new member of the KRE2/MNT1 family.";
RL Yeast 10:819-831(1994);
RN [2]
RP SEQUENCE OF 726-747, AND CHARACTERIZATION.
RC STRAIN=Y57;
RX MEDLINE=95021683; PubMed=7935765;
RA Reese J.C., Apone L., Walker S.S., Griffin L.A., Green M.R.;

```

```

RT "Yeast TAFIIS in a multisubunit complex required for activated
RT transcription.";
RT Nature 371:523-527(1994).
RN [3]
RP SEQUENCE OF 22-63 AND 726-752, AND CHARACTERIZATION.
RC STRAIN=YPH252;
RX MEDLINE=95396770; PubMed=7667272;
RA Poon D., Bai Y., Campbell A.M., Bjorklund S., Kim Y.-J., Zhou S.,
RA Kornberg R.D., Well P.A.;
RT "Identification and characterization of a TFIID-like multiprotein
RT complex from Saccharomyces cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:8224-8228(1995).
CC -1- FUNCTION: TAFS ARE COMPONENTS OF THE TRANSCRIPTION FACTOR IID
CC (TFIID) COMPLEX THAT ARE ESSENTIAL FOR MEDIATING REGULATION OF
CC RNA POLYMERASE TRANSCRIPTION.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs) WHOSE MW RANGE FROM
CC 25-150 kDa.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTRAINS 5 WD REPEATS (TBP-ASP DOMAINS).
CC -1- SIMILARITY: TO DROSOPHILA TAFII-80 AND MAMMALIAN TAF2D.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: 221487; CAA79685.1; -
DR DDB: 236067; ZAB85160.1; -
DR PIR: S34023; S34023.
DR SGI: S0000402; TAF90.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40.6.
DR PRINTS: PR00320; GPROTEINBRT.
DR SMART: SM00320; WD40.6.
DR PROSITE: PS00678; WD_REPEATS_1; 3.
DR PROSITE: PS50082; WD_REPEATS_2; 6.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Nuclear protein; Transcription regulation; Repeat; WD repeat.
FT REPEAT 523 553 WD 1.
FT REPEAT 565 595 WD 2.
FT REPEAT 607 637 WD 3.
FT REPEAT 649 679 WD 4.
FT REPEAT 691 722 WD 5.
SQ SEQUENCE 798 AA; 88967 MW; B42315B8C752D0B6 CRC64;

Query Match 9.5%; Score 188; DB 1; Length 798;
Best Local Similarity 20.4%; Pred. No. 4.3e-08;
Matches 65; Conservative 64; Mismatches 110; Indels 80; Gaps 12;

OY 74 LOSYADEDEKESESYTYSMACGVNGNRYVAGGVKGIIRVIDVNSEIHH----- 121
DB 459 MYTFQNTNMDSCLDPSDDCRI-----AAAGFDOSYIKIWSLDSGLNPNALNNNDX 513
OY 122 -----KSLVGHGDSVNEIRPOLKPOLVITASKDESVRLNVTGICILIFAGAGHRE 176
DB 514 EDPTCKTIVLGHSGTYVSTFSP-DNKYLLSGSDKTVRLMSDTHALYSK---GHNP 569
OY 237 SIHTNVVDCNRMF--GDFILKSVDNEILLMEPOLKENSPEGASDVLLRYPPKDI-- 292
DB 606 AGLHNDVDCVSHPNMGCVYFTGSSDCTCRMV-----STGDSVR-LFGLHAPVSTIV 659
OY 293 -----WFIKSCDLHLSVAIGNOGKRYVWDLKSCPVLTIKLSHNSKSVYRTTAN 347
DB 660 CPDGRV-----LSTGSEDCIINWMDIGTOKRL-----KQMGHCKNAIYSLSYSK 704

```



DR SMART; SM00320; WD40; 7.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 3.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 6.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Transcription regulation; Repressor; Repeat; WD repeat.  
 FT REPEAT 285 325 WD 1.  
 FT REPEAT 332 371 WD 2.  
 FT REPEAT 374 413 WD 3.  
 FT REPEAT 415 454 WD 4.  
 FT REPEAT 456 495 WD 5.  
 FT REPEAT 510 549 WD 6.  
 FT REPEAT 552 585 WD 7.  
 SQ SEQUENCE 586 AA; 64192 MW; 79C991ABDDC7D095 CRC64;

Query Match 9.3%; Score 183.5; DB 1; Length 586;  
 Best Local Similarity 22.1%; Pred. No. 6.9e-08;  
 Matches 87; Conservative 62; Mismatches 161; Indels 83; Gaps 17;

QY 7 GNESTVGLTSPN-----KSKYVTNRIOEGKKPLYAVNFPLD-----ARFP--DV 51  
 DB 241 GNPPTPAELIIPSNVNRREKDMVTSTNPNKEPPIVOLHLEHTSVICVRFESADK 300  
 QY 52 FYTAGGNRTILYNCIGDAISALQSYADEDKESFYTVSWACGVNGNPVYAGVGIIIR 111  
 DB 301 FLATGNGNRAAMFNVEYTGKLTLLQESSKREGDLVRSVAFSPDCK-VLATGEEDQOIR 359  
 QY 112 VIDVNSETHKSLVGHGDSVNEIRPOLKPOLYITASKDESVRLMNVETGICILIRGAG 111  
 DB 360 INDIAOKRVRRLTGHQOEIYSLDFSK-DGKTLVSGSGDRCTVCLMVEAGEOKLILHTD 418  
 QY 172 GHRVEVLSVDFHPSDIYRASCMDTTIKIMSKKEFWYVEKSFMTDPSKPTKPTVYOF 231  
 DB 419 G-----VTTVMFSPDQOF-IAAGSLDKVIRIWTSS--GTLVEQLHGHEES----- 460  
 QY 232 PVFTASIRHNYVDCNRMFGDFILSKSVNDEILMEPOLKEN--SPG-----EGA 278  
 DB 461 -YYSVAFSPD-----GKYLVSGLDMNTIKMLQCVSNVAPSMYKKEGICKQTFEGH 511  
 QY 279 SDVLLRPPPMCDIWFIRKSCDLHLSSVALIGNQGVYVWDLKSCPPVLITKLSHNSOKS 338  
 DB 512 KPEILSVTVSPDGKWTIS-----GSKRTIIOFMSPPS-PSHSLTGLGHNNSS-- 556  
 QY 339 VIRQTAMSYDSTILACCEDGT-----IWRMD 365  
 DB 557 -VISAVASPNGH-----CFATSGDLRLRIMWYE 584

RESULT 9  
 ID PF20\_CHLRE STANDARD; PRT; 606 AA.  
 AC P93107;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Flagellar WD-repeat protein PF20.  
 GN PF20.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith E., Lefevre P.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: INTER-MICROTUBULE BRIDGES IN FLAGELLA.  
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.ch](mailto:license@isb.ch)).

DR EMBL; U78547; AABA1727.1; -  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 7.  
 DR PRINTS; PR00320; GPTEINBRPT.  
 DR SMART; SM00320; GPTEINBRPT.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 4.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 6.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat; Flagella.  
 FT REPEAT 324 354 WD 1.  
 FT REPEAT 366 396 WD 2.  
 FT REPEAT 408 438 WD 3.  
 FT REPEAT 450 480 WD 4.  
 FT REPEAT 492 522 WD 5.  
 FT REPEAT 534 564 WD 6.  
 FT REPEAT 576 606 WD 7.  
 SQ SEQUENCE 606 AA; 65839 MW; 7A8779501E3218D1 CRC64;

Query Match 9.2%; Score 182; DB 1; Length 606;  
 Best Local Similarity 23.5%; Pred. No. 9.6e-08;  
 Matches 87; Conservative 57; Mismatches 134; Indels 92; Gaps 19;

QY 5 TLGNESVGLTSPNKKSYVTNRIOEGKKPLYAVNFPLDARFDFVYTAGGNRTILYN 64  
 DB 236 SLGGQSTAAAGGAGSAGALCATNRALTDVPPAA-----GAAAAAATGR----- 279  
 QY 65 CLDGDAISALQSYADEDKESFYTVSWACGVNG-----NPVYAGVGVIIRVIDVSETI 120  
 DB 280 -----SGAVSAP-----RSGNA-SLNAPRRNPADLEFPAPAKMLSLN----- 319  
 QY 121 HKSIVGHGDSVNEIRPOLKPOLYITASKDESVRLMNVETGICILIFAGAGHREYELV 180  
 DB 320 -KTFKGHLISVANILAHPTKP-ILVYASDDKTKMMHMPGG--DLIMCGE-GHKDWVAGV 374  
 QY 181 DFHPSDIYRASCMDTTIKIMSKKEFWYVEK-----FTTDDPSKPTKPTVYFPVFTAS 237  
 DB 375 DFHPAGTC-LASGGDSAVKIMDF-----EKQRCVTTPTDHKQ-----AIVSVR 417  
 QY 238 IHTNIVYDCNRMFGDFILSKSVNDEILMEPOLKENSPEGASVLLKYPVMCDIWFIRK 297  
 DB 418 FH-----HLGEVVASGSLDHTVRLMD-----LPAGCRNALRGHVDVNDLMDQPF 463  
 QY 298 SCDLHLSSVALIGNQGVYVWDLKS--CPVLITKLSHNSOKSVIROTAMSYDSTILAC 355  
 DB 464 S-----SLATASSDKTVSWDARAGICTG---TYGHQNSCGV---SPNIGTQLAST 512  
 QY 356 CEDGTIWRMD 365  
 DB 513 DADGVVXKLM 522

RESULT 10  
 ID GBB\_MAIZE STANDARD; PRT; 380 AA.  
 AC P49178;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Guanine nucleotide-binding protein beta subunit.  
 GN GBI.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Lillipsidea; Poales; Poaceae; PACC clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95023950; PubMed=7937804;  
 RA Weiss C.A., Garnaat C.W., Mukai K., Hu Y., Ma H.;

RT \*Isolation of cDNAs encoding guanine nucleotide-binding protein beta-subunit homologues from maize (ZGB1) and Arabidopsis (AGB1)."

RT Proc. Natl. Acad. Sci. U.S.A. 91:9554-9558(1994).

CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN- CC EFFECTOR INTERACTION.

CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).

CC -1- TISSUE SPECIFICITY: PRESENT IN THE ROOT, LEAF, AND TASSSEL.

CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration CC between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: U12233; AAA50446.1; -

DR HSSP: P04901; 1TBG.

DR MaizeDB: 105669; -

DR InterPro: IPR001632; Gprotein\_B.

DR InterPro: IPR001680; WD40.

DR Pfam: PF00400; WD40; 7.

DR PRINTS: PR00320; GPROTEINB.

DR SMART: SM00320; WD40; 7.

DR PROSITE: PS00678; WD\_REPEATS\_1; 3.

DR PROSITE: PS50082; WD\_REPEATS\_2; 5.

DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.

DR Transducer; Repeat; WD repeat.

FT REPEAT 64 94 WD 1.

FT REPEAT 106 136 WD 2.

FT REPEAT 155 186 WD 3.

FT REPEAT 203 234 WD 4.

FT REPEAT 247 277 WD 5.

FT REPEAT 296 326 WD 6.

FT REPEAT 342 372 WD 7.

FT SEQUENCE 380 AA; 41714 MW; 7D5DC5276C33D1E CRC64;

QY Query Match 9.2%; Score 181.5; DB 1; Length 380;

QY Best Local Similarity 24.9%; Pred. No. 5,7e-08;

QY Matches 80; Conservative 46; Mismatches 142; Indels 53; Gaps 12;

QY 59 RITLYNCLGDAISALOSYADEDEKESFYVSMACGVNPNPYAAGYKGIIRIVDNSE 118

QY 89 RLIVWNLTSQKTHAIKHP-----WVACAFAPNGOS-VACGGIDSACISIFNLNSQ 140

QY 119 T-----IHSLYGHGDSVNEIRTPKLPOLVTASDESVRLMNVETGICLIIFAG-- 169

QY 141 ADRGNMNPVSRILGHHGVSSCOYVPQDETRLLITSSGDCVLMVDTTGRIISIFGGEF 200

QY 170 AGHREYVLSDYFHSPIYRASCMDTTIKIMSKFEWTVESFTWDDPSKFPTEFV 229

QY 201 PSGRADVQSVSSINSMFMVSGSCDTTVRLMDIRISRAVRYTHGHEDVNS--VAF- 257

QY 230 QPVTFTASHTNYVDCNFMFGDFILSKSVNELLIMEP-----OLKENSPEGASDVLL 283

QY 258 -FP-----DGNR-FG-----TGSDDGTCRLFDWRTGHQLOVYSREDDRNSE--- 297

QY 284 RYVPMDIMWIKFSCDLHLSVAAGNEGKYVWDLSKSPVLLTKLSHNSKSVINQOT 343

QY 298 -----LPTVTSIAFSISGR-----LFAGYSGDCGYWDTLLAEVVLNGLNLSHGRISCL 350

QY 344 AMSVSGSTIIACCEDEGTIRW 364

QY 351 GMSSDGSAICTGSMWKIKIW 371

Db

RESULT 11

TRCB\_XENIA STANDARD; PRT; 518 AA.

ID TRCB\_XENIA

AC Q91854; P70037; P70038;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Beta-Tricp (beta-transducin repeat-containing protein).

GN FBXW1 OR BTFCP.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;

OC Xenopodidae; Xenopus.

OX NCBI\_TaxID=8353;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9330289; PubMed=8393141;

RA Szevak W., Keiper B.D., Stratowa C., Castanon M.J.;

RT "Saccharomyces cerevisiae cdc15 mutants arrested at a late stage in anaphase are rescued by Xenopus cDNAs encoding N-ras or a protein with beta-transducin repeats."

RT Mol. Cell. Biol. 13:4953-4966(1993).

RL [2]

RP SEQUENCE OF 302-518 FROM N.A.

RX MEDLINE=97109804; PubMed=8952061;

RA Hudson J.W., Alarcon V.B., Elinson R.P.;

RT "Identification of new localized RNAs in the Xenopus oocyte by RT differential display PCR."

RL Dev. Genet. 19:190-198(1996).

CC -1- FUNCTION: Probably recognizes and binds to some phosphorylated proteins and promotes their ubiquitination and degradation.

CC -1- SUBUNIT: PART OF A SCF (SKP1-CUL1IN-F-BOX) PROTEIN LIGASE COMPLEX (BY SIMILARITY).

CC -1- DEVELOPMENTAL STAGE: PRESENT IN FULLY GROWN AND PROGESTERONE-MATURED OOCYTES. THE LEVEL CHANGE VERY LITTLE EVEN AFTER ZYGOTIC GENE TRANSCRIPTION BEGINS FOLLOWING THE MIDBLASTULA TRANSITION. DO NOT INCREASE IN ABUNDANCE IN THE GASTRULA, NEURULA, TAILBUD, OR TADPOLE EMBRIO.

CC -1- SIMILARITY: CONTAINS 1 F-BOX DOMAIN.

CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).

CC -1- SIMILARITY: STRONG, TO C.ELEGANS K10B2.1.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration CC between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M96268; AAA02810.1; -

DR EMBL: U63921; AAB49671.1; -

DR EMBL: U63922; AAB49672.1; -

DR InterPro: IPR001810; F-box.

DR InterPro: IPR001680; WD40.

DR Pfam: PF00646; F-box; 1.

DR Pfam: PF00400; WD40; 7.

DR PRINTS: PR00320; GPROTEINBPT.

DR SMART: SM00256; FBOX; 1.

DR SMART: SM00320; WD40; 7.

DR PROSITE: PS50181; FBOX; 1.

DR PROSITE: PS00678; WD\_REPEATS\_1; 6.

DR PROSITE: PS50082; WD\_REPEATS\_2; 7.

DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.

DR Ubiquitin conjugation; Repeat; F-box.

FT DOMAIN 119 157

FT REPEAT 230 258 WD 1.

FT REPEAT 270 298 WD 2.

FT REPEAT 310 338 WD 3.

FT REPEAT 353 381 WD 4.

FT REPEAT 393 421 WD 5.

FT REPEAT 433 461 WD 6.

FT REPEAT 482 510 WD 7.

FT CONFLICT 302 304 GEM -> EFR (IN REF. 2).

FT CONFLICT 516 518 GIA -> AAH (IN REF. 2).  
SQ SEQUENCE 518 AA: 59507 MW: 2A52EC19028127F3 CRC64;

Query Match 9.1%; Score 179.5; DB 1; Length 518;  
Best Local Similarity 25.4%; Pred. No. 1.3e-07;  
Matches 75; Conservative 45; Mismatches 108; Indels 67; Gaps 14;

QY 87 YTSWACGVNAGNPNVAAAGVKGKIIIVDVSSETHKSLVGHGSDVNEIRTPQLPOLVIT 146  
DB 271 HIGSVLCLOYDERVITITGSSDSTVRWVDVNTGEMLNTLHHCBAVLHLR--FNNGMAMV 327  
QY 147 ASKDESRLMNVETGICILIFAGAGHRYEVLSDPHSPDIYFASCGMDTIKISMKR 206  
DB 328 CSKDRSIAMVDMASTPDIPLRLVYGHRAAVNVDF--DDKIYVSAAG--DRTIKVMN--- 381  
QY 207 FWTYVEKSTWTDPSKFPKVFQPFVFTASHTNTVYDCNRMGDFILSKSVNELLIME 266  
DB 382 -----TSTCEF-VRTLNGHKRGIACTQYDRLVVSGSSDNTIRLMD 421  
QY 267 POLKENSPP-----EGASDVLRLRYVPMCDIWFIFSCDLHSSVAIGNQEGKYVWML 320  
DB 422 IEC-----GACLRVLEGHELV-----RC-----IRFD-----NKRIVSAGAYDKIKYVMDL 462  
QY 321 KSC-----PPV-----LITKLSHNSQSKSVIRQTAMSVDSGTILACCEGDTIWRMDVI 367  
DB 463 VALDPRAPAGTLCRLTYEH--SGRVFR---LQDFEQLVSSSHDITILIMDFL 512

## RESULT 12

FM1A\_HUMAN STANDARD: PRT: 605 AA.

AC 09Y297: 09Y213;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE F-box/WD-repeat protein 1B (F-box and WD-repeats protein beta-TrCP)  
DE (E3SISIKRAB) (pikappaB) (pikappaB) (E3 receptor subunit).  
GN FBXK1A OR FBXK1A OR BTBRC OR BTBRC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=99075339; PubMed=9859996;  
RA Yaron A., Hatzubai A., Davis M., Lavon I., Amit S., Manning A.M.,  
RT Andersen J.S., Mann M., Mercurio F., Ben-Neriah Y.;  
RT "Identification of the receptor component of the IkarappaBalpha-  
ubiquitin ligase";  
RL Nature 396:590-594(1998).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Lymphoid;  
RX MEDLINE=98325370; PubMed=9660940;  
RA Margottin F., Bour S.P., Durand H., Selig L., Benichou S., Richard V.,  
RT Thomas D., Strebel K., Benarous R.;  
RT "A novel human WD protein, h-beta TrCP, that interacts with HIV-1 Ypu  
RT connects CD4 to the ER degradation pathway through an F-box motif";  
RL Mol. Cell 1:565-574(1998).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RX MEDLINE=20003060; PubMed=10531035;  
RA Centaurelli C., Chlaur D.S., Guadavaccaro D., Parks W., Vidal M.,  
RT Pagano M.;  
RT "Identification of a family of human F-box proteins";  
RL Curr. Biol. 9:1177-1179(1999).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=99145464; PubMed=9990852;  
RA Winston J.T., Strack P., Beer-Romero P., Chu C.Y., Elledge S.J.,  
RA Harper J.W.;  
RT "The SCF(beta-TRCP)-ubiquitin ligase complex associates specifically

RT with phosphorylated destruction motifs in I-kappa-B-alpha and  
RT beta-catenin and stimulates I-kappa-B-alpha ubiquitination in vitro.";  
RL Genes Dev. 13:270-283(1999)  
CC - FUNCTION: BINDS SPECIFICALLY TO PHOSPHORYLATED IKB ALPHA  
CC (PIKAPPAALPHA) AND PHOSPHORYLATED BETA-CATENIN AND PROMOTES THEIR  
CC UBIQUITINATION AND DEGRADATION.  
CC - SUBUNIT: PART OF A SCF (SKP1-CUL1N-F-BOX) PROTEIN LIGASE COMPLEX.  
CC - SUBCELLULAR LOCATION: Cytoplasmic.  
CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC - SIMILARITY: CONTAINS 1 F-BOX DOMAIN.  
CC - SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: AF101784; AAD08702.1;  
DR EMBL: YL4153; CAA74572.1;  
DR EMBL: AF129530; AAF04464.1;  
DR MIM: 603482;  
DR InterPro: IPR001810; F-box.  
DR InterPro: IPR001680; WD40.  
DR Pfam: PF00646; F-box; 1.  
DR Pfam: PF00400; WD40; 7.  
DR PRINTS: PR00320; GPROTEINBPT.  
DR SMART: SM00256; FBOX; 1.  
DR SMART: SM00320; WD40; 7.  
DR PROSITE: PS50181; FBOX; 1.  
DR PROSITE: PS00678; WD\_REPEATS\_1; 6.  
DR PROSITE: PS50082; WD\_REPEATS\_2; 7.  
DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
KW Ubiquitin conjugation; Repeat; WD repeat; Alternative splicing.  
FT DOMAIN 190 228  
FT REPEAT 301 338  
FT REPEAT 341 378  
FT REPEAT 381 418  
FT REPEAT 424 461  
FT REPEAT 464 503  
FT REPEAT 505 541  
FT REPEAT 553 590  
FT REPEAT 590 605  
FT VARSPLIC 17 52  
SQ SEQUENCE 605 AA: 68866 MW: 4C67F3B7E400ED37 CRC64;

Query Match 9.0%; Score 177.5; DB 1; Length 605;  
Best Local Similarity 25.4%; Pred. No. 2.3e-07;  
Matches 75; Conservative 45; Mismatches 108; Indels 67; Gaps 14;

QY 87 YTSWACGVNAGNPNVAAAGVKGKIIIVDVSSETHKSLVGHGSDVNEIRTPQLPOLVIT 146  
DB 342 HTSVLCLOYDERVITITGSSDSTVRWVDVNTGEMLNTLHHCBAVLHLR--FNNGMAMV 398  
QY 147 ASKDESRLMNVETGICILIFAGAGHRYEVLSDPHSPDIYFASCGMDTIKISMKR 206  
DB 399 CSKDRSIAMVDMASTPDIPLRLVYGHRAAVNVDF--DDKIYVSAAG--DRTIKVMN--- 452  
QY 207 FWTYVEKSTWTDPSKFPKVFQPFVFTASHTNTVYDCNRMGDFILSKSVNELLIME 266  
DB 453 -----TSTCEF-VRTLNGHKRGIACTQYDRLVVSGSSDNTIRLMD 492  
QY 267 POLKENSPP-----EGASDVLRLRYVPMCDIWFIFSCDLHSSVAIGNQEGKYVWML 320  
DB 493 IEC-----GACLRVLEGHELV-----RC-----IRFD-----NKRIVSAGAYDKIKYVMDL 533  
QY 321 KSC-----PPV-----LITKLSHNSQSKSVIRQTAMSVDSGTILACCEGDTIWRMDVI 367  
DB 534 VALDPRAPAGTLCRLTYEH--SGRVFR---LQDFEQLVSSSHDITILIMDFL 583



```

RESULT 13
GBL_SOLUTION
ID GBL_SOLUTION STANDARD: PRT; 377 AA.
AC P93563;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit.
GN GBL.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OC NCBI_TaxID=4113;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. DESIREE;
RA Provart N.J., Ma H., Willmitzer L., Mueller-Roeber B.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEIN (G PROTEIN) ARE
CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE
CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE
CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-
CC EFFECTOR INTERACTION.
CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X87837; CAA61106.1; -
CC HSSP; P04901; 17BG.
CC InterPro: IPR001632; Gproteins_B.
CC InterPro: IPR001680; WD40.
CC Pfam; PF00400; WD40; 7.
CC PRINTS; PR00319; GPROTEINB.
CC SMART; SM00320; GPROTEINBPT.
CC PROSITE; PS00678; WD_REPEATS_1; 2.
CC PROSITE; PS00682; WD_REPEATS_2; 5.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC Transducer; Repeat; WD repeat.
CC FT REPEAT 63 93 WD 1.
CC FT REPEAT 105 135 WD 2.
CC FT REPEAT 154 185 WD 3.
CC FT REPEAT 202 233 WD 4.
CC FT REPEAT 246 276 WD 5.
CC FT REPEAT 293 323 WD 6.
CC FT REPEAT 339 369 WD 7.
CC FT REPEAT 377 399 WD 7.
CC SEQUENCE 377 AA; 40941 MW; B9AFLAAG72IDCBCL CRC64;
Query Match 8.9%; Score 176; DB 1; Length 377;
Best Local Similarity 23.2%; Pred. No. 1.6e-07;
Matches 86; Conservative 49; Mismatches 156; Indels 80; Gaps 14;
QY 30 QEGKKPLAAVFNFLD-----ARFDFVETAGNRRTLYNCGDAISALQSYA 78
DB 42 KQGSNP---VFEGPTDLVCCRLGHTGKYSIDMTPEKNRIV--SASODGLIYWNALIT 96
QY 79 DEDKESEFYTYSWACGVGNP---VVAAGVYGIIRVIDVNS---ETIH---KSLVGHG 128
DB 97 SQKTAHAIKPCAMWMTCAFSPSGQSVACGGLDSACSIFNLNPSIDKGIHPYSRLMGGHK 156
QY 129 DSVNEIRIPLKPOLVITASKDESRLKNNVETGICILIFAG--AGCHREVLVSYPFHSD 186
DB 157 GYVSSCQYVPEDTHLITSSGDQTCVLDITTGRTSVFGEFGSGHTRADVLVSISSSN 216

```

```

QY 187 IYRFASCGMDTIKINSMKEEFWYVERKSTWTDPSKFKYQFPVFTASIHITNYDCN 246
DB 217 PKLFVSGSCDPTARLMD-----TRVASRAQRTFGHESDVNFKF-----PDDGN 261
QY 247 RWFQDFILSKSVNDEILME-----POLKENSPEGASDYLRRPVPKCDIMPIKFSODLH 302
DB 262 R-FG---TSSDQSCSRLFIIRGHQLYVNOPHGDDI-----PH 297
QY 303 LSSVA-----IGNQEGKVVYMDLKSPPYLITKLISHNOSKSVIRGTAMSVDSSTIL 353
DB 298 VTSMAEFISGRILEFVGSNDCYVWDLLAKVYLNIGSVQNSHGRISCLGLSADGSALC 357
QY 354 ACCEDGTIWRM 364
DB 358 TGSWDTNLKIM 368
RESULT 14
PKWA_THECU
ID PKWA_THECU STANDARD: PRT; 742 AA.
AC P49655;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative serine/threonine-protein kinase pkwa (EC 2.7.1.-).
DE PKWA OR PKW1.
GN Thermomonospora curvata.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Thermomonosporaceae;
OC Thermomonospora.
OX NCBI_TaxID=2020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCM 3352;
RX MEDLINE=96200125; PubMed=8631732;
RA Janda L., Tichy P., Spizke J., Petricek M.;
RT "A deduced Thermomonospora curvata protein containing
RT serine/threonine protein kinase and WD-repeat domains."
RL J. Bacteriol. 178:1487-1489(1996).
CC -1- FUNCTION: MAY PLAY A REGULATORY ROLE DURING THE COMPLEX GROWTH
CC CYCLE AND IN SECONDARY METABOLITE PRODUCTION.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF115313; AAB05822.1; -
CC HSSP; P00523; 2PTK.
CC InterPro: IPR000719; Euk_Pkinase.
CC InterPro: IPR002290; Ser_thr_Pkinase.
CC InterPro: IPR001680; WD40.
CC Pfam; PF00069; Pkinase; 1.
CC Pfam; PF00400; WD40; 7.
CC PRINTS; PR00320; GPROTEINBPT.
CC SMART; SM00320; WD40; 7.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS00678; WD_REPEATS_1; 5.
CC PROSITE; PS00682; WD_REPEATS_2; 7.
CC PROSITE; PS50082; WD_REPEATS_REGION; 1.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC Transferrase; Serine/threonine-protein kinase; ATP-binding; Repeat;
KW WD repeat.
KW DOMAIN 16 266 PROTEIN KINASE.
KW NP_BIND 22 30 ATP (BY SIMILARITY).
KW BINDING 44 44 ATP (BY SIMILARITY).

```



Query Match	8.9%	Score 175.5	DB 1	Length 1146
Best Local Similarity	22.4%	Pred. No. 8e-07		
Matches 89	Conservative 55	Mismatches 136	Indels 115	Gaps 20
5	TLGNESIVGSLPNSKKSKYKVTNRIOEGKKPLVAVVFNLDPARFVDVETAGNRITLYN	64		
Db	823 TWMVDPIDLELMPD-----NTIKVGAKQLPKRAEFSKKDLK-----	860		
QY	65 CLGDGAISALQSYADEDEKEESFYTVSMACGVGNPNVAAAGYKGIIRVIDNSE-----T	119		
Db	861 -----CVSTIQSFRRRVNSIAFF-----DNOKILLCAGYDGTGYRVFVDVDMNKCLYT	907		
QY	120 IHSKSLVGHDSVNEILRTQPLKPOLVITFASKDSVRLMNVENG--ICIIIFAGAGGHREY	177		
Db	908 VN-----GHHKSIESTI---ACNSNNTYFTSSPDNTIVHIIIRSGNTKCIETLV---GHTGEV	957		
QY	178 LSVDEHPSDIYFFASCQMDTTIKISMKKEFWYVE-KSFVTWDDPSKEPTKEVQEPVFTA	236		
Db	956 NCV--VANENKYL-F-SCSYDKITIKVMDLS---TFKEIKSF-----E	991		
QY	237 SIHTNYVDNCNRMFGDFTILSKSVDNELLMEPO---LKENSPE-----GASDVILR	284		

